PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C07K 07/64, 07/10, A61K 39/21
A61K 39/395, C07K 15/28

A2

(11) International Publication Number: WO 91/15512

(43) International Publication Date: 17 October 1991 (17.10.91)

(21) International Application Number: PCT/US91/02166

(22) International Filing Date: 1 April 1991 (01.04.91)

(30) Priority data:
504,772 3 April 1990 (03.04.90) US

(71) Applicant: GENENTECH, INC. [US/US]; 460 Point San Bruno Boulevard, South San Francisco, CA 94080 (US).

(72) Inventors: GREGORY, Timothy, J. ; 414 Pinehill Road, Hillsborough, CA 94010 (US). LEONARD, Cordelia, K. ; 923 Crespi Drive, Pacifica, CA 94044 (US). SPELLMAN, Michael, W. ; 468 Oak Avenue, San Bruno, CA 94066 (US).

(74) Agents: ADLER, Carolyn, R. et al.; Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

**Published**
*Without international search report and to be republished upon receipt of that report.*

(54) Title: HIV ENVELOPE POLYPEPTIDES

(57) Abstract

Novel isolated physiologically active polypeptides are provided, as well as antibodies directed against the isolated polypeptides. Methods for the preparation and pharmaceutical use of the polypeptides and antibodies are also provided.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | |
|---|---|---|---|---|---|
| AT | Austria | ES | Spain | MG | Madagascar |
| AU | Australia | FI | Finland | ML | Mali |
| BB | Barbados | FR | France | MN | Mongolia |
| BE | Belgium | GA | Gabon | MR | Mauritania |
| BF | Burkina Faso | GB | United Kingdom | MW | Malawi |
| BG | Bulgaria | GN | Guinea | NL | Netherlands |
| BJ | Benin | GR | Greece | NO | Norway |
| BR | Brazil | HU | Hungary | PL | Poland |
| CA | Canada | IT | Italy | RO | Romania |
| CF | Central African Republic | JP | Japan | SD | Sudan |
| CG | Congo | KP | Democratic People's Republic of Korea | SE | Sweden |
| CH | Switzerland | | | SN | Senegal |
| CI | Côte d'Ivoire | KR | Republic of Korea | SU | Soviet Union |
| CM | Cameroon | LI | Liechtenstein | TD | Chad |
| CS | Czechoslovakia | LK | Sri Lanka | TG | Togo |
| DE | Germany | LU | Luxembourg | US | United States of America |
| DK | Denmark | MC | Monaco | | |

The major envelope glycoprotein (gp120) of HIV-1 has been the object of intensive investigation since the initial identification of HIV-1 as the etiological agent of AIDS (Barre-Sinoussi *et al.*, 1983). The gp120 molecule is of interest as a vaccine candidate (Berman *et al.*, 1988; Arthur *et al.*, 1987), as the mediator of viral attachment via the virus receptor CD4 (Dalgleish *et al.*, 1984; Klatzman *et al.*, 1984) and the spread of the virus by cell-to-cell fusion (syncytia formation), and as an agent with immunosuppressive effects of its own (Shalaby *et al.*, 1987; Diamond *et al.*, 1988). It is also a potential mediator of the pathogenesis of HIV-1 in AIDS (Siliciano *et al.*, 1988; Sodroski *et al.*, 1986) and has been suggested to be the viral protein most accessible to immune attack.

Currently, gp120 is considered to be the best candidate for a subunit vaccine, because: (i) gp120 is known to possess the CD4 binding domain by which HIV attaches to its target cells, (ii) HIV infectivity can be neutralized *in vitro* by antibodies to gp 120, (iii) the majority of the *in vitro* neutralizing activity present in the serum of HIV infected individuals can be removed with a gp120 affinity column, and (iv) the gp120/gp41 complex appears to b essential for the transmission of HIV by cell-to-cell fusion. See, e.g. Hu *et al.*, *Nature* 328:721-724 (1987) (vaccinia virus-HIV env recombinant vaccine); Arthur *et al.*, *J. Virol.* 63(12): 5046-5053 (1989) (purified gp120); and Berman *et al.*, *Proc. Natl. Acad. Sci. USA* 85:5200-5204 (1988) (recombinant envelope glycoprotein gp120).

The gp120 molecule is synthesized as part of a membrane-bound glycoprotein, gp160 (Allan *et al.*, 1985). Via a host-cell mediated process, gp160 is cleaved to form gp120 and the integral membrane protein gp41 (Robey *et al.*, 1985). Together gp120 and gp41 form the spikes observed on the surface of newly released HIV-1 virions (Gelderblom *et al.*, 1987). As there is no covalent attachment between gp120 and gp41, free gp120 is released from the surface of virions and infected cells (Gelderblom *et al.*, 1985).

The gp120 molecule consists of a polypeptide core of 60,000 daltons; extensive modification by N-linked glycosylation increases the apparent molecular weight of the molecule to 120,000 (Lasky *et al., Science*, 233:209-212 (1986)). The amino acid sequence of gp120 contains five relatively conserved domains interspersed with five hypervariable domains (Modrow *et al., J. Virology* 61(2):570 (1987); Willey *et al., Proc. Natl. Acad. Sci. USA* 83:5038-5042 (1986)). The hypervariable domains contain extensive amino acid substitutions, insertions and deletions. Sequence variations in these domains result in up to 25% overall sequence variability between gp120 molecules from the various viral isolates. Despite this variation, several structural and functional elements of gp120 are highly conserved. Among these are the ability of gp120 to bind to the viral receptor CD4, the ability of gp120 to interact with gp41 to induce fusion of the viral and host cell membranes, the positions of the 18 cysteine residues in the gp120 primary sequence, and the positions of 13 of the approximately 24 N-linked glycosylation sites in the gp120 sequence.

# HIV ENVELOPE POLYPEPTIDES

## Field of the Invention

This invention is concerned with antigens of the HIV virus, and to novel physiologically active polypeptides found in the HIV env glycoprotein.

## Background of the Invention

Acquired immunodeficiency syndrome (AIDS) is caused by a retrovirus identified as the human immunodeficiency virus (HIV). A number of immunologic abnormalities have been described in AIDS including abnormalities in B-cell function, abnormal antibody response, defective monocyte cell function, impaired cytokine production, depressed natural killer and cytotoxic cell function, and defective ability of lymphocytes to recognize and respond to soluble antigens. Other immunologic abnormalities associated with AIDS have been reported. Among the more important immunologic defects in patients with AIDS is the depletion of the T4 helper/inducer lymphocyte population.

In spite of the profound immunodeficiency observed in AIDS, the mechanism(s) responsible for immunodeficiency are not clearly understood. Several postulates exist. One accepted view is that defects in immune responsiveness are due to selective infection of helper T cells by HIV resulting in impairment of helper T-cell function and eventual depletion of cells necessary for a normal immune response. *In vitro* and *in vivo* studies showed that HIV can also infect monocytes which are known to play an essential role as accessory cells in the immune response. HIV may also result in immunodeficiency by interfering with normal cytokine production in an infected cell resulting in secondary immunodeficiency as for example, IL-1 and IL-2 deficiency. An additional means of HIV-induced immunodeficiency consists of the production of factors which are capable of suppressing the immune response. None of these models resolves the question of whether a component of HIV *per se*, rather than infection by replicative virus, is responsible for the immunologic abnormalities associated with AIDS.

The HIV env protein has been extensively described, and the amino acid and RNA sequences encoding HIV env from a number of HIV strains are known (Modrow, S. *et al., J. Virology* 61(2): 570 (1987). The HIV virion is covered by a membrane or envelope derived from the outer membrane of host cells. The membrane contains a population of envelope glycoproteins (gp 160) anchored in the membrane bilayer at their carboxyl terminal region. Each glycoprotein contains two segments. The N-terminal segment, called gp120 by virtue of its relative molecular weight of about 120kD, protrudes into the aqueous environment surrounding the virion. The C-terminal segment, called gp41, spans the membrane. gp120 and gp 41 are linked by a peptide bond that is particularly susceptible to proteolytic cleavage, see e.g. McCune *et al., EPO Application No. 0 335 635*, priority 28 March 88 and references cited therein.

Many workers in the field have prepared mutagenic and fragment variants of gp120. See, e.g.: Matsushita *et al., J. Virology* 62:2107-2114 (1988); Rusche *et al., Proc. Natl. Acad. Sci. USA* 85:3198-3202 (1988); Goudsmit *et al., AIDS* 2:157-164 (1988); Javaherian *et al., Proc. Natl. Acad. Sci. USA* 86:6768-6772 (1989); Lasky *et al., Cell* 50:975-985 (1987); Kowalski *et al., Science* 237:1351-1355 (1987); Willey *et al., Proc. Natl. Acad. Sci. USA* 83;5038-5042 (1986); Modrow *et al., J. Virology* 61:570-578 (1987).

The disulfide bonding pattern within gp120 and the positions of actual oligosaccharide moieties on the molecule would be useful information for directing mutagenesis and fragmentation studies aimed at defining the functional domains of gp120 and sites for potential pharmacological interruption of its functions (e.g., type-common neutralizing epitopes). This information has been difficult to obtain due to the small amounts of gp120 available from natural sources, the complexity of the disulfide bonding and oligosaccharide structures in gp120, and uncertainty regarding the functionality or structural relevance (Moore *et al.*, in press) of rgp120 produced in non-mammalian systems.

The inventors herein have surprisingly discovered that certain regions of native gp120 exist in specific three-dimensional conformation, which conformation is conserved over isotype and strain.

It is an object of this invention to provide novel polypeptides which are useful as diagnostic tools for assaying biological samples for evidence of HIV infection.

It is a further object of this invention to provide novel polypeptides which are usable for vaccines, and for pharmacologic interruption of the course of HIV infection.

It is a further object of this invention to provide methods for preparing such polypeptides, and antibodies directed to such polypeptides.

Other objects, features, and characteristics of the present invention will become apparent upon consideration of the following description and the appended claims.

Summary of the Invention

The objects of this invention are accomplished by the preparation and administration of compositions comprising isolated cyclized polypeptides which are suitable for administration to a human or non-human patient having or at risk of having HIV infection. These cyclized polypeptides are selected from the following:

a) C V K L T P L C C N T S V I T Q A C [SEQ. ID NO. 1] and containing less than about 28 amino acid residues;

b) P I H Y C A P A G F A I L K C N N K T F N G T G P C T N V S T V Q C T H G I R P [SEQ. ID NO. 2] and containing less than about 45 amino acid residues;

c) C N N K T F N G T G P C [SEQ. ID NO. 3] and containing less than about 22 amino acid residues;

d) C A P A G F A I L K C C T N V S T V Q C [SEQ. ID NO. 4] and containing less than about 30 amino acid residues;

e) P I H Y C C T H G I R P [SEQ. ID NO. 5] and containing less than about 22 amino acid residues;

f) G G D P E I V T H S F N C G G E F F Y C N S L P C R I K Q F I N M W Q E V G K A M Y A P P I S G Q I R C S S N I T G [SEQ. ID. NO. 6] and containing less than about 65 amino acid residues;

g) C G G E F F Y C C R I K Q F I N M W Q E V G K A M Y A P P I S G Q I R C [SEQ. ID NO. 7] and containing less than about 45 amino acid residues;

h) C A S D A K A Y D T E V H N V W A T H A C [SEQ. ID NO. 8] and containing less than about 30 amino acid residues; and

i) T T T L F C A S D A K A Y D T E V H N V W A T H A C V P T D P N [SEQ. ID NO. 9] and containing less than about 50 amino acid residues.

Additionally, this invention is also directed to compositions comprising an isolated polypeptide having an antigenic determinant or determinants immunologically cross-reactive with a determinant of the HIV env polypeptide of strain HTLV-IIIB having an amino acid sequence selected from the group consisting of

a) residues 1-80;

b) residues 8-180;

c) residues 165-260;

d) residues 160-260;

e) residues 260-310; and

f) residues 320-479.

This invention is particularly directed to vaccines comprising the compositions of this invention. The compositions of this invention, including variant analogues thereof, are also useful in diagnostic assays for HIV neutralizing antibody in patient samples.

Monoclonal antibodies directed to the isolated polypeptides of this invention are provided, characterized by their affinity for ligand, epitope binding, and ability to a) block CD4/gp120 binding, b) neutralize HIV virions, c) reduce reverse transcriptase activity *in vitro*, and d) inhibit syncytia formation.

These antibodies are useful as diagnostics for the presence of HIV infection in a patient or patient sample, and for affinity purification of HIV env. These antibodies are also useful in passively immunizing patients infected with HIV. In certain embodiments, antibodies are provided which are conjugated to a cytotoxin, a water-insoluble matrix, or to a detectable marker.

Antibodies directed to HIV env epitopes have been described in the literature; however, it should be noted that, due to the variety and confusion among authors currently as to numbering systems for HIV env sequences, not all antibodies described in the literature as directed to certain regions will actually the same residue numbers as defined herein (see e.g. Matsushita *et al.*, *J. Virol.* 62:2107-2114 (1988); EPO Application No. EP 339 504; Rusche

*et al., Proc. Natl. Acad. Sci. USA*, 85:3198-3202 (1988); Looney *et al., Science* 241:357-359 (1988);

Brief Description of the Drawings

FIGURE 1 provides the amino acid sequences of (a) the mature envelope glycoprotein (gp120) from the $III_B$ isolate of HIV-1 [SEQ. ID NO. 10], and (b) the N-terminal sequence portion of the recombinant fusion glycoproteins (9AA [SEQ. ID NO. 11] or CL44 [SEQ. ID NO. 12]) from the herpes simplex gD1. Fusion sites between the gD1 and gp120 segments in the 9AA and CL44 constructions are marked with (°) and (°°), respectively. The letter T refers to observed tryptic cleavage of the gp120 segment, and the peptides are ordered sequentially starting at the N-terminus of the molecule. Lower case letters following the T number indicate other unexpected proteolytic cleavages. The letter H refers to the observed tryptic cleavage of the herpes simplex gD1 protein portion of CL44. Peptide T2' contains the fusion site in CL44. The cysteine residues of gp120 are shaded, and potential N-glycosylation sites are indicated with a dot above the corresponding asparagine residue.

FIGURE 2 shows a reversed-phase HPLC tryptic map of RCM CL44. This chromatogram was generated with 7.5 nmol of trypsin-digested RCM CL44. Chromatography conditions were as described in Experimental Procedures. Peaks were collected and identified by AAA and in some cases confirmed by N-terminal sequence analysis (Table I). Identified peaks are labelled according to the nomenclature given in Figure 1. Peptides containing potential tryptic sites that were not hydrolyzed are designated by two T numbers separated by a comma.

FIGURE 3 shows a reversed-phase HPLC tryptic map of 9AA. This chromatogram was generated with 6.8 nmol of sample. Chromatography conditions were as described in the Example herein. Peaks containing cysteine residues were identified by N-terminal sequence analysis. These identifications are summarized in Table II.

FIGURE 4 shows the results of further manipulations of tryptic peptides from the map of 9AA to isolate individual disulfides. The chromatograms are details of microbore reversed-phase HPLC separations of peptides resulting from: (a) treatment of peptides T12, T13, and T14 (Peak C, Figure 3) with PNGase F followed by endoproteinase Asp-N, (b) treatment of peptides T3, T4, and T11 (Peak F, Figure 3) with PNGase F followed by endoproteinase Asp-N, and (c) treatment of peptides T28 and T31 (Peak D, Figure 3) with *S. aureus* V8 protease. Chromatography conditions were as described in the Example herein. Peak identifications were determined by N-terminal sequence analysis and are given in Table III.

FIGURE 5 shows reverse-phase HPLC tryptic maps of endoglycosidase treated RCM CL44. The chromatograms are tryptic maps of: (a) untreated RCM CL44, (b) PNGase

F-treated RCM CL44, and (c) endo H-treated RCM CL44. Each tryptic map was generated with 7.5 nmol of sample. Chromatography conditions were as described in Experimental Procedures. Peaks were collected and identified by AAA (data not shown). Glycopeptide peaks are labelled according to the nomenclature in Figure 1.

FIGURE 6 is a schematic representation of gp 120 of the $III_B$ isolate of HIV-1 showing disulfides and glycosylation sites, with the amino acids represented in single-letter code [SEQ. ID NO. 10]. Roman numerals label the five disulfide-bonded domains. The five hypervariable regions of Modrow *et al.*, *J. Virol.* 61:570-578 (1987) are enclosed in boxes and labelled V1-V5. Glycosylation sites containing high mannose-type and/or hybrid-type oligosaccharide structures are indicated by a branching-Y symbol, and glycosylation sites containing complex-type oligosaccharide structures are indicated by a V-shaped symbol.

FIGURE 7 shows a schematic representation of the HIV env glycoprotein gp120 of HIV-2, showing disulfides and potential glycosylation sites [SEQ. ID NO. 13]. Glycosylation sites are indicated by a shaded box around a N residue. Roman numerals label five disulfide-bonded domains.

Detailed Description of the Invention

HIV env is defined herein as the envelope polypeptide of Human Immunodeficiency Virus as described above, together with its amino acid sequence variants and derivatives produced by covalent modification of HIV env or its variants *in vitro*, as discussed herein. As used herein, the term "HIV env" encompasses all forms of gp120 and/or 160, e.g. including fragments, fusions of gp160/120 or their fragments with other peptides, and variantly glycosylated or unglycosylated HIV env. The HIV env of this invention is recovered free of active virus.

HIV env and its variants are conventionally prepared in recombinant cell culture. For example, see EP publication No. 187041. Henceforth, gp120 prepared in recombinant cell culture is referred to as rgp120. Recombinant synthesis is preferred for reasons of safety and economy, but it is known to prepare peptides by chemical synthesis and to purify HIV env from viral culture; such env preparations are included within the definition of HIV env herein.

Genes encoding HIV env are obtained from the genomic cDNA of an HIV strain or from available subgenomic clones containing the gene encoding HIV env.

This invention is directed to isolated polypeptides. Certain of these isolated polypeptides are defined as cyclized polypeptides comprising a particular amino acid sequence, and certain isolated polypeptides are described by reference to specific amino acid residue numbers. The amino acid numbering reflects the mature HIV-1 gp120 amino acid sequence as shown by Fig. 6. and Fig. 1A [SEQ. ID NO. 10], not counting any signal sequence or other upstream regions, and is used throughout this description to conveniently connote the intended residues, however it is understood that this invention is not limited to

those specific residue numbers. For gp120 sequences which include the native HIV-IIIB N-terminal signal sequence, numbering may differ. The same nucleotide and amino acid residue numbers may not be applicable in other strains where upstream deletions or insertions change the length of the viral genome and HIV env, but the region encoding this portion of gp120 is readily identified by reference to the teachings herein. Also, variant signal sequences (such as those resulting from a fusion with a fragmented or heterologous signal sequence as discussed below may lead to a slightly different numbering, however the precise amino acid sequences are discerned for all embodiments by reference to Fig. 6 and/or Fig 1A [SEQ. ID NO. 10].

Included within the scope of the isolated polypeptides of this invention, as those terms are used herein are polypeptides having specified amino acid sequences, deglycosylated or unglycosylated derivatives, homologous amino acid sequence variants, and homologous *in vitro*-generated variants and derivatives, and which variants are capable of exhibiting a biological activity in common with the HIV env of Fig. 6 or Fig. 7.

Isolated polypeptide biological activity is defined as either 1) immunological cross-reactivity with at least one isolated polypeptide, or 2) the possession of at least one adhesive or effector function qualitatively in common with the isolated polypeptide. Examples of the qualitative biological activities of an isolated polypeptide include the ability to bind to the viral receptor CD4 or known monoclonal antibodies, and the ability of gp120 to interact with gp41 to induce fusion of the viral and host cell membranes.

Immunologically cross-reactive as used herein means that the candidate polypeptide is capable of competitively inhibiting the qualitative biological activity of an isolated polypeptide having this activity with polyclonal antisera raised against the known active analogue. Such antisera are prepared in conventional fashion by injecting goats or rabbits, for example, subcutaneously with the known active analogue in complete Freund's adjuvant, followed by booster intraperitoneal or subcutaneous injection in incomplete Freunds.

The ordinarily skilled worker may use the disulfide bonding pattern within gp120 and the positions of actual oligosaccharide moieties on the molecule as described herein for directing mutagenesis and fragmentation variants of the claimed isolated polypeptides. It is intended that the variants of this invention include isolated polypeptides in which one or more residues have been substituted, deletions of one or more residues, and insertions of one or more amino acid residues.

This invention also contemplates amino acid sequence variants of the isolated polypeptides. Amino acid sequence variants are prepared with various objectives in mind, including increasing the affinity of the isolated polypeptide for a ligand or antibody, facilitating the stability, purification and preparation of the isolated polypeptide, modifying its plasma half life, improving therapeutic efficacy, and lessening the severity or occurrence of side effects during therapeutic use of the isolated polypeptide. In the discussion below, amino acid

sequence variants of the isolated polypeptide are provided, exemplary of the variants that may be selected.

Amino acid sequence variants of isolated polypeptide fall into one or more of three classes: Insertional, substitutional, or deletional variants. These variants ordinarily are prepared by site-specific mutagenesis of nucleotides in the DNA encoding the isolated polypeptide, by which DNA encoding the variant is obtained, and thereafter expressing the DNA in recombinant cell culture. However, fragments having up to about 100-150 amino acid residues are prepared conveniently by *in vitro* synthesis. The following discussion applies to any isolated polypeptide to the extent it is applicable to its structure or function.

The amino acid sequence variants of the isolated polypeptide are predetermined variants not found in nature or naturally occurring alleles. The isolated polypeptide variants typically exhibit the same qualitative biological--for example, antibody binding--activity as the naturally occurring isolated polypeptide or isolated polypeptide analogue. However, isolated polypeptide variants and derivatives that are not capable of binding to antibodies are useful nonetheless (a) as a reagent in diagnostic assays for isolated polypeptide or antibodies to the isolated polypeptide, (b) when insolubilized in accord with known methods, as agents for purifying anti-isolated polypeptide antibodies from antisera or hybridoma culture supernatants, and (c) as immunogens for raising antibodies to isolated polypeptide or as immunoassay kit components (labelled, as a competitive reagent for the native isolated polypeptide or unlabelled as a standard for isolated polypeptide assay) so long as at least one isolated polypeptide epitope remains active.

While the site for introducing an amino acid sequence variation is predetermined, the mutation *per se* need not be predetermined. For example, in order to optimize the performance of a mutation at a given site, random or saturation mutagenesis (where all 20 possible residues are inserted) is conducted at the target codon and the expressed isolated polypeptide variant is screened for the optimal combination of desired activities. Such screening is within the ordinary skill in the art.

Amino acid insertions usually will be on the order of about from 1 to 10 amino acid residues; substitutions are typically introduced for single residues; and deletions will range about from 1 to 30 residues. Deletions or insertions preferably are made in adjacent pairs, i.e. a deletion of 2 residues or insertion of 2 residues. It will be amply apparent from the following discussion that substitutions, deletions, insertions or any combination thereof are introduced or combined to arrive at a final construct. Insertional amino acid sequence variants of the isolated polypeptide are those in which one or more amino acid residues extraneous to the isolated polypeptide are introduced into a predetermined site in the target isolated polypeptide and which displace the preexisting residues.

Commonly, insertional variants are fusions of heterologous proteins or polypeptides to the amino or carboxyl terminus of the isolated polypeptide. Such variants are referred to as

fusions of the isolated polypeptide and a polypeptide containing a sequence which is other than that which is normally found in the isolated polypeptide at the inserted position. Several groups of fusions are contemplated herein.

The novel isolated polypeptides of this invention are useful in diagnostics or in purification of the antibodies or ligands by known immunoaffinity techniques.

Desirable fusions of the isolated polypeptide, which may or may not also be immunologically active, include fusions of the mature isolated polypeptide sequence with a signal sequence heterologous to a native isolated polypeptide as mentioned above. Signal sequence fusions are employed in order to more expeditiously direct the secretion of the isolated polypeptide. The heterologous signal replaces the native isolated polypeptide signal, and when the resulting fusion is recognized, i.e. processed and cleaved by the host cell, the isolated polypeptide is secreted. Signals are selected based on the intended host cell, and may include bacterial yeast, mammalian and viral sequences. The native HIV env signal or the herpes gD glycoprotein signal is suitable for use in mammalian expression systems.

C-terminal or N-terminal fusions of the isolated polypeptide or isolated polypeptide fragment with an immunogenic hapten or heterologous polypeptide are useful as vaccine components for the immunization of patients against HIV infection. Fusions of the hapten or heterologous polypeptide with isolated polypeptide or its active fragments which retain T-cell binding activity are also useful in directing cytotoxic T cells against target cells where the hapten or heterologous polypeptide is capable of binding to a target cell surface receptor.

The precise site at which the fusion is made is variable; particular isolated polypeptide sites are selected in order to optimize the biological activity, secretion or binding characteristics of the isolated polypeptide. The optimal site will for a particular application will be determined by routine experimentation.

Substitutional variants are those in which at least one residue in the isolated polypeptide has been removed and a different residue inserted in its place. Such substitutions generally are made in accordance with the following Table 1 when it is desired to finely modulate the characteristics of the isolated polypeptide.

TABLE 1

| Original Residue | Exemplary Substitutions |
|---|---|
| Ala | ser |
| Arg | lys |
| Asn | gln; his |
| Asp | glu |
| Cys | ser; ala |
| Gln | asn |
| Glu | asp |
| Gly | pro |
| His | asn; gln |
| Ile | leu; val |
| Leu | ile; val |
| Lys | arg; gln; glu |
| Met | leu; ile |
| Phe | met; leu; tyr |
| Ser | thr |
| Thr | ser |
| Trp | tyr |
| Tyr | trp; phe |
| Val | ile; leu |

Novel amino acid sequences, as well as isosteric analogs (amino acid or otherwise), as included within the scope of this invention.

Substantial changes in function or immunological identity are made by selecting substitutions that are less conservative than those in Table 1, i.e., selecting residues that differ more significantly in their effect on maintaining (a) the structure of the polypeptide backbone in the area of the substitution, for example as a sheet or helical conformation, (b) the charge or hydrophobicity of the molecule at the target site or (c) the bulk of the side chain. The substitutions which in general are expected to produce the greatest changes in isolated polypeptide properties will be those in which (a) a hydrophilic residue, e.g. seryl or threonyl, is substituted for (or by) a hydrophobic residue, e.g. leucyl, isoleucyl, phenylalanyl, valyl or alanyl; (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain, e.g., lysyl, arginyl, or histidyl, is substituted for (or by) an electronegative residue, e.g., glutamyl or aspartyl; or (d) a residue having a bulky side chain, e.g., phenylalanine, is substituted for (or by) one not having a side chain, e.g., glycine.

Some deletions, insertions, and substitutions will not produce radical changes in the characteristics of the isolated polypeptide molecule. However, when it is difficult to predict

the exact effect of the substitution, deletion, or insertion in advance of doing so, for example when modifying an immune epitope, one skilled in the art will appreciate that the effect will be evaluated by routine screening assays. For example, a variant typically is made by site specific mutagenesis of the isolated polypeptide -encoding nucleic acid, expression of the variant nucleic acid in recombinant cell culture and, optionally, purification from the cell culture for example by immunoaffinity adsorption on a polyclonal anti-isolated polypeptide column (in order to adsorb the variant by at least one remaining immune epitope). The activity of the cell lysate or purified isolated polypeptide variant is then screened in a suitable screening assay for the desired characteristic. For example, a change in the immunological character of the isolated polypeptide, such as affinity for T-cell binding, is measured by a competitive-type immunoassay. As more becomes known about the functions *in vivo* of the isolated polypeptide other assays will become useful in such screening. Modifications of such protein properties as redox or thermal stability, hydrophobicity, susceptibility to proteolytic degradation, or the tendency to aggregate with carriers or into multimers are assayed by methods well known to the artisan.

Another class of isolated polypeptide variants are deletional variants. Deletions are characterized by the removal of one or more amino acid residues from the isolated polypeptide sequence. Typically, deletions are used to affect isolated polypeptide biological activities, however, deletions which preserve the biological activity or immune cross-reactivity of the isolated polypeptide are suitable.

Deletions of cysteine or other labile residues also may be desirable, for example in increasing the oxidative stability of the isolated polypeptide. Deletion or substitutions of potential proteolysis sites, e.g. Arg Arg, is accomplished by deleting one of the basic residues or substituting one by glutaminyl or histidyl residues.

It will be understood that some variants may exhibit reduced or absent biological activity. These variants nonetheless are useful as standards in immunoassays for the isolated polypeptide so long as they retain at least one immune epitope of the isolated polypeptide.

It is presently believed that the three-dimensional structure of the isolated polypeptides and peptide compositions of the present invention is important to their functioning as described herein. Therefore, all related structural analogs which mimic the active structure of those formed by the isolated polypeptides claimed herein are specifically included within the scope of the present invention.

Glycosylation variants are included within the scope of the isolated polypeptide. They include variants completely lacking in glycosylation (unglycosylated) and variants having at least one less glycosylated site than the native form (deglycosylated) as well as variants in which the glycosylation has been changed. Included are deglycosylated and unglycosylated amino acid sequence variants, deglycosylated and unglycosylated isolated polypeptide having the native, unmodified amino acid sequence of the isolated polypeptide, and other

glycosylation variants. For example, substitutional or deletional mutagenesis is employed to eliminate the N- or O-linked glycosylation sites of the isolated polypeptide, e.g., an asparagine residue (not at the clip site) is deleted or substituted for by another basic residue such as lysine or histidine. Alternatively, flanking residues making up the glycosylation site are substituted or deleted, even though the asparagine residues remain unchanged, in order to prevent glycosylation by eliminating the glycosylation recognition site.

Unglycosylated isolated polypeptide which has the amino acid sequence of the native isolated polypeptide is produced in recombinant prokaryotic cell culture because prokaryotes are incapable of introducing glycosylation into polypeptides.

Glycosylation variants are produced by selecting appropriate host cells or by *in vitro* methods. Yeast, for example, introduce glycosylation which varies significantly from that of mammalian systems. Similarly, mammalian cells having a different species (e.g. hamster, murine, insect, porcine, bovine or ovine) or tissue origin (e.g. lung, liver, lymphoid, mesenchymal or epidermal) than the source of the isolated polypeptide antigen are routinely screened for the ability to introduce variant glycosylation as characterized for example by elevated levels of mannose or variant ratios of mannose, fucose, sialic acid, and other sugars typically found in mammalian glycoproteins. *In vitro* processing of the isolated polypeptide typically is accomplished by enzymatic hydrolysis, e.g. neuraminidase digestion.

Covalent modifications of the isolated polypeptide molecule which do not modify the clip site are included within the scope hereof. Such modifications are introduced by reacting targeted amino acid residues of the recovered protein with an organic derivatizing agent that is capable of reacting with selected side chains or terminal residues, or by harnessing mechanisms of post-translational modification that function in selected recombinant host cells. The resulting covalent derivatives are useful in programs directed at identifying residues important for biological activity, for immunoassays of isolated polypeptide or for the preparation of anti-isolated polypeptide antibodies for immunoaffinity purification of the recombinant isolated polypeptide. For example, complete inactivation of the biological activity of the protein after reaction with ninhydrin would suggest that at least one arginyl or lysyl residue is critical for its activity, whereafter the individual residues which were modified under the conditions selected are identified by isolation of a peptide fragment containing the modified amino acid residue. Such modifications are within the ordinary skill in the art and are performed without undue experimentation.

Derivatization with bifunctional agents is useful for preparing intermolecular aggregates of the isolated polypeptide with polypeptides as well as for cross-linking the isolated polypeptide to a water insoluble support matrix or surface for use in the assay or affinity purification of its ligands. In addition, a study of intrachain cross-links will provide direct information on conformational structure. Commonly used cross-linking agents include sulfhydryl reagents, 1,1-bis(diazoacetyl)-2-phenylethane, glutaraldehyde, N-

hydroxysuccinimide esters, for example esters with 4-azidosalicylic acid, homobifunctional imidoesters including disuccinimidyl esters such as 3,3'-dithiobis (succinimidyl-propionate), and bifunctional maleimides such as bis-N-maleimido-1,8-octane. Derivatizing agents such as methyl-3-[(p-azido-phenyl)dithio] propioimidate yield photoactivatable intermediates which are capable of forming cross-links in the presence of light. Alternatively, reactive water insoluble matrices such as cyanogen bromide activated carbohydrates and the systems reactive substrates described in U.S. patents 3,959,080; 3,969,287; 3,691,016; 4,195,128; 4,247,642; 4,229,537; 4,055,635; and 4,330,440 are employed for protein immobilization and cross-linking.

Polymers generally are covalently linked to the isolated polypeptide herein through a multifunctional crosslinking agent which reacts with the polymer and one or more amino acid or sugar residues of protein. However, it is within the scope of this invention to directly crosslink the polymer by reacting a derivatized polymer with the isolated polypeptide, or vice versa. Covalent bonding to amino groups is accomplished by known chemistries based upon cyanuric chloride, carbonyl diimidazole, aldehyde reactive groups (PEG alkoxide plus diethyl acetal of bromoacetaldehyde; PEG plus DMSO and acetic anhydride, or PEG chloride plus the phenoxide of 4-hydroxybenzaldehyde, succinimidyl active esters, activated dithiocarbonate PEG, 2,4,5-trichlorophenylchloroformate or p-nitrophenylchloroformate activated PEG. Carboxyl groups are derivatized by coupling PEG-amine using carbodiimide.

This invention is also directed to polypeptides of this invention which by definition or optionally are conformationally stabilized by cyclization. The peptides ordinarily are cyclized by covalently bonding the N and C-terminal domains of one peptide to the corresponding domain of another peptide of this invention so as to form cyclooligomers containing two or more iterated peptide sequences, each internal peptide having substantially the same sequence. Further, cyclized peptides (whether cyclooligomers or cylomonomers) are crosslinked to form 1-3 cyclic structures having from 2 to 6 peptides comprised therein. The peptides preferably are not covalently bonded through $\alpha$-amino and -carboxyl groups (head to tail), but rather are cross-linked through the side chains of residues located in the N and C-terminal domains. The linking sites thus generally will be between the side chains of $A_1$ and $A_{10}$ residues. Substantially identical polypeptides present in the polymerized forms of the peptides hereof are those which exhibit qualitative isolated polypeptide activity, notwithstanding the degree of amino acid sequence variation among the polypeptides. Variants which exhibit activity are used as subunits in homo or heteropolymers. In homopolymers the peptides are the same. Heteropolymers contain different peptides, each however, chosen from within the parameters described above.

Many suitable methods *per se* are known for preparing mono- or poly-cyclized peptides as contemplated herein. Lys/Asp cyclization has been accomplished using N$\alpha$-Boc-amino

acids on solid-phase support with Fmoc/OFm side-chain protection for Lys/Asp; the process is completed by piperidine treatment followed by BoP cyclization.

Glu and Lys side chains also have been crosslinked in preparing cyclic or bicyclic peptides: the peptide is synthesized by solid phase chemistry on a p-methylbenzhydrylamine resin. The peptide is cleaved from the resin and deprotected. The cyclic peptide is formed using diphenylphosphorylazide in dilute dimethylformamide. For an alternative procedure, see Schiller *et al.*, "Peptide Protein Res." 25:171-177 (1985). See also U.S. Patent 4,547,489.

Disulfide crosslinked or cyclized peptides are generated by conventional methods. The method of Pelton *et al.* (J. Med. Chem. 29:2370-2375 [1986]) is suitable, except that a greater proportion of cyclooligomers are produced by conducting the reaction in more concentrated solutions than the dilute reaction mixture described by Pelton *et al.* for the production of cyclomonomers. The same chemistry is useful for synthesis of dimers (using $A_1$-$A_9$ Pen plus $A_1$-$A_9$ Cys) or cyclooligomers or cyclomonomers (Pen $A_1$-$A_{10}$ Cys, or Pen $A_1$-$A_{10}$ Cys plus Cys $A_1$-$A_{10}$ Pen). Also useful are thiomethylene bridges (Tetrahedron Letters 25(20):2067-2068 [1984]). See also Cody *et al.*, J. Med. Chem. 28:583 (1985).

The desired cyclic or polymeric peptides are purified by gel filtration followed by reversed-phase high pressure liquids chromatography or other conventional procedures. The peptides are sterile filtered and formulated into conventional pharmacologically acceptable vehicles.

Certain post-translational derivatizations are the result of the action of recombinant host cells on the expressed polypeptide. Glutaminyl and asparaginyl residues are frequently post-translationally deamidated to the corresponding glutamyl and aspartyl residues. Alternatively, these residues are deamidated under mildly acidic conditions. Either form of these residues falls within the scope of this invention.

Other post-translational modifications include hydroxylation of proline and lysine, phosphorylation of hydroxyl groups of seryl or threonyl residues, methylation of the $\alpha$-amino groups of lysine, arginine, and histidine side chains (T.E. Creighton, Proteins: Structure and Molecular Properties, W.H. Freeman & Co., San Francisco pp 79-86 [1983]), acetylation of the N-terminal amine and, in some instances, amidation of the C-terminal carboxyl.

DNA encoding the isolated polypeptide is synthesized by *in vitro* methods or is obtained readily from cDNA libraries. The means for synthetic creation of the DNA encoding the isolated polypeptide, either by hand or with an automated apparatus, are generally known to one of ordinary skill in the art, particularly in light of the teachings contained herein. As examples of the current state of the art relating to polynucleotide synthesis, one is directed to Maniatis *et al.*, *Molecular Cloning--A Laboratory Manual*, Cold Spring Harbor Laboratory (1984), and Horvath *et al.*, *An Automated DNA Synthesizer Employing Deoxynucleoside 3'-Phosphoramidites*, Methods in Enzymology 154: 313-326, 1987.

Alternatively, to obtain DNA encoding the isolated polypeptide, one needs only to conduct hybridization screening with labelled DNA encoding either the isolated polypeptide or isolated polypeptide fragment (usually, greater than about 20, and ordinarily about 50bp) in order to detect clones which contain homologous sequences in the cDNA libraries derived from cells or tissues of a particular animal, followed by analyzing the clones by restriction enzyme analysis and nucleic acid sequencing to identify full-length clones. If full length clones are not present in the library, then appropriate fragments are recovered from the various clones and ligated at restriction sites common to the fragments to assemble a full-length clone. DNA encoding isolated polypeptide from various isotypes and strains is obtained by probing libraries from hosts of such species with the amino acid sequences of the isolated polypeptide, or by synthesizing the genes *in vitro*.

In general, prokaryotes are used for cloning of DNA sequences in constructing the vectors useful in the invention. For example, *E. coli* K12 strain 294 (ATCC No. 31446) is particularly useful. Other microbial strains which may be used include *E. coli* B and *E. coli* X1776 (ATCC No. 31537). These examples are illustrative rather than limiting. Alternatively, *in vitro* methods of cloning, e.g. polymerase chain reaction, are suitable.

The isolated polypeptides of this invention are expressed directly in recombinant cell culture as an N-terminal methionyl analogue, or as a fusion with a polypeptide heterologous to the hybrid/portion, preferably a signal sequence or other polypeptide having a specific cleavage site at the N-terminus of the hybrid/portion. For example, in constructing a prokaryotic secretory expression vector for the isolated polypeptide, the native isolated polypeptide signal is employed with hosts that recognize that signal. When the secretory leader is "recognized" by the host, the host signal peptidase is capable of cleaving a fusion of the leader polypeptide fused at its C-terminus to the desired mature isolated polypeptide. For host prokaryotes that do not process the native isolated polypeptide signal, the signal is substituted by a prokaryotic signal selected for example from the group of the alkaline phosphatase, penicillinase, lpp or heat stable enterotoxin II leaders. For yeast secretion the native isolated polypeptide signal may be substituted by the yeast invertase, alpha factor or acid phosphatase leaders. In mammalian cell expression the native isolated polypeptide signal or native HIV <u>env</u> signal is satisfactory for certain isolated polypeptides, although other mammalian secretory protein signals are suitable, as are viral secretory leaders, for example the herpes simplex gD signal.

The isolated polypeptide may be expressed in any host cell, but preferably is synthesized in mammalian hosts. However, host cells from prokaryotes, fungi, yeast, insects and the like are also are used for expression. Exemplary prokaryotes are the strains suitable for cloning as well as *E. coli* W3110 ($F^-\lambda^-$ prototrophic, ATTC No. 27325), other

enterobacteriaceae such as *Serratia marcescans*, bacilli and various pseudomonads. Preferably the host cell should secrete minimal amounts of proteolytic enzymes.

Expression hosts typically are transformed with DNA encoding the isolated polypeptide which has been ligated into an expression vector. Such vectors ordinarily carry a replication site (although this is not necessary where chromosomal integration will occur). Expression vectors also include marker sequences which are capable of providing phenotypic selection in transformed cells, as will be discussed further below. For example, *E. coli* is typically transformed using pBR322, a plasmid derived from an *E. coli* species (Bolivar, *et al.*, Gene 2: 95 [1977]). pBR322 contains genes for ampicillin and tetracycline resistance and thus provides easy means for identifying transformed cells, whether for purposes of cloning or expression. Expression vectors also optimally will contain sequences which are useful for the control of transcription and translation, e.g., promoters and Shine-Dalgarno sequences (for prokaryotes) or promoters and enhancers (for mammalian cells). The promoters may be, but need not be, inducible; even powerful constitutive promoters such as the CMV promoter for mammalian hosts may produce the isolated polypeptide without host cell toxicity. While it is conceivable that expression vectors need not contain any expression control, replicative sequences or selection genes, their absence may hamper the identification of transformants and the achievement of high level peptide expression.

Promoters suitable for use with prokaryotic hosts illustratively include the β-lactamase and lactose promoter systems (Chang *et al.*, *Nature* 275: 615 [1978]; and Goeddel *et al.*, *Nature* 281: 544 [1979]), alkaline phosphatase, the tryptophan (trp) promoter system (Goeddel, *Nucleic Acids Res.* 8: 4057 (1980) and EPO Appln. Publ. No. 36,776) and hybrid promoters such as the tac promoter (H. de Boer *et al.*, *Proc. Natl. Acad. Sci. USA* 80: 21-25 [1983]). However, other functional bacterial promoters are suitable. Their nucleotide sequences are generally known, thereby enabling a skilled worker operably to ligate them to DNA encoding the isolated polypeptide (Siebenlist *et al.*, *Cell* 20: 269 [1980]) using linkers or adaptors to supply any required restriction sites. Promoters for use in bacterial systems also will contain a Shine-Dalgarno (S.D.) sequence operably linked to the DNA encoding the isolated polypeptide.

In addition to prokaryotes, eukaryotic microbes such as yeast or filamentous fungi are satisfactory. *Saccharomyces cerevisiae* is the most commonly used eukaryotic microorganism, although a number of other strains are commonly available. The plasmid YRp7 is a satisfactory expression vector in yeast (Stinchcomb, *et al.*, *Nature* 282: 39 (1979); Kingsman *et al*, *Gene* 7: 141 (1979); Tschemper *et al.*, *Gene* 10: 157 (1980)). This plasmid already contains the trp1 gene which provides a selection marker for a mutant strain of yeast lacking the ability to grow in tryptophan, for example ATCC no. 44076 or PEP4-1 (Jones, *Genetics* 85: 12 [1977]). The presence of the trp1 lesion as a characteristic of the yeast

host cell genome then provides an effective environment for detecting transformation by growth in the absence of tryptophan.

Suitable promoting sequences for use with yeast hosts include the promoters for 3-phosphoglycerate kinase (Hitzeman *et al., J. Biol. Chem.* 255: 2073 (1980)) or other glycolytic enzymes (Hess *et al., J. Adv. Enzyme Reg.* 7: 149 (1968); and Holland, *Biochemistry* 17: 4900 (1978)), such as enolase, glyceraldehyde-3-phosphate dehydrogenase, hexokinase, pyruvate decarboxylase, phosphofructokinase, glucose-6-phosphate isomerase, 3-phosphoglycerate mutase, pyruvate kinase, triosephosphate isomerase, phosphoglucose isomerase, and glucokinase.

Other yeast promoters, which are inducible promoters having the additional advantage of transcription controlled by growth conditions, are the promoter regions for alcohol dehydrogenase 2, isocytochrome C, acid phosphatase, degradative enzymes associated with nitrogen metabolism, metallothionein, glyceraldehyde-3-phosphate dehydrogenase, and enzymes responsible for maltose and galactose utilization. Suitable vectors and promoters for use in yeast expression are further described in R. Hitzeman *et al.*, European Patent Publication No. 73,657A.

Expression control sequences are known for eucaryotes. Virtually all eukaryotic genes have an AT-rich region located approximately 25 to 30 bases upstream from the site where transcription is initiated. Another sequence found 70 to 80 bases upstream from the start of transcription of many genes is a CXCAAT region where X may be any nucleotide. At the 3' end of most eukaryotic genes is an AATAAA sequence which may be the signal for addition of the poly A tail to the 3' end of the coding sequence. All of these sequences are inserted into mammalian expression vectors.

Suitable promoters for controlling transcription from vectors in mammalian host cells are readily obtained from various sources, for example, the genomes of viruses such as polyoma virus, SV40, adenovirus, MMV (steroid inducible), retroviruses (e.g. the LTR of HIV), hepatitis-B virus and most preferably cytomegalovirus, or from heterologous mammalian promoters, e.g. the beta actin promoter. The early and late promoters of SV40 are conveniently obtained as an SV40 restriction fragment which also contains the SV40 viral origin of replication. Fiers *et al., Nature*, 273: 113 (1978). The immediate early promoter of the human cytomegalovirus is conveniently obtained as a HindIII E restriction fragment. Greenaway, P.J. *et al., Gene* 18: 355-360 (1982).

Transcription of a DNA encoding the isolated polypeptide by higher eukaryotes is increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10-300bp, that act on a promoter to increase its transcription. Enhancers are relatively orientation and position independent having been found 5' (Laimins *et al., PNAS* 78: 993 [1981]) and 3' (Lusky, M.L., *et al., Mol. Cell Bio.* 3: 1108 (1983)) to the transcription unit, within an intron (Banerji, J.L. *et al., Cell* 33: 729

(1983)) as well as within the coding sequence itself (Osborne, T.F., *et al., Mol. Cell Bio.* 4: 1293 [1984]). Many enhancer sequences are now known from mammalian genes (globin, elastase, albumin, α-fetoprotein and insulin). Typically, however, one will use an enhancer from a eukaryotic cell virus. Examples include the SV40 enhancer on the late side of the replication origin (bp 100-270), the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

Expression vectors used in eukaryotic host cells (yeast, fungi, insect, plant, animal, human or nucleated cells from other multicellular organisms) will also contain sequences necessary for the termination of transcription which may affect mRNA expression. These regions are transcribed as polyadenylated segments in the untranslated portion of the mRNA encoding the hybrid immunoglobulin. The 3' untranslated regions also include transcription termination sites.

Expression vectors may contain a selection gene, also termed a selectable marker. Examples of suitable selectable markers for mammalian cells are dihydrofolate reductase (DHFR), thymidine kinase (TK) or neomycin. When such selectable markers are successfully transferred into a mammalian host cell, the transformed mammalian host cell is able to survive if placed under selective pressure. There are two widely used distinct categories of selective regimes. The first category is based on a cell's metabolism and the use of a mutant cell line which lacks the ability to grow independent of a supplemented media. Two examples are CHO DHFR$^-$ cells and mouse LTK$^-$ cells. These cells lack the ability to grow without the addition of such nutrients as thymidine or hypoxanthine. Because these cells lack certain genes necessary for a complete nucleotide synthesis pathway, they cannot survive unless the missing nucleotides are provided in a supplemented media. An alternative to supplementing the media is to introduce an intact DHFR or TK gene into cells lacking the respective genes, thus altering their growth requirements. Individual cells which were not transformed with the DHFR or TK gene will not be capable of survival in non-supplemented media. In preferred embodiments, herein, CHO cells which are DHFR$^-$ are used for recombinant expression of the isolated polypeptide.

The second category of selective regimes is dominant selection which refers to a selection scheme used in any cell type and does not require the use of a mutant cell line. These schemes typically use a drug to arrest growth of a host cell. Those cells which are successfully transformed with a heterologous gene express a protein conferring drug resistance and thus survive the selection regimen. Examples of such dominant selection use the drugs neomycin (Southern *et al., J. Molec. Appl. Genet.* 1: 327 (1982)), mycophenolic acid (Mulligan *et al., Science* 209: 1422 (1980)) or hygromycin (Sugden *et al., Mol. Cell. Biol.* 5: 410-413 (1985)). The three examples given above employ bacterial genes under eukaryotic control to convey resistance to the appropriate drug G418 or neomycin (geneticin), xgpt (mycophenolic acid) or hygromycin, respectively.

"Amplification" refers to the increase or replication of an isolated region within a cell's chromosomal DNA. Amplification is achieved using a selection agent, e.g. methotrexate (MTX) which inactivates DHFR. Amplification or the making of successive copies of the DHFR gene results in greater amounts of DHFR being produced in the face of greater amounts of MTX. Amplification pressure is applied notwithstanding the presence of endogenous DHFR, by adding ever greater amounts of MTX to the media. Amplification of a desired gene can be achieved by cotransfecting a mammalian host cell with a plasmid having a DNA encoding a desired protein and the DHFR or amplification gene permitting cointegration. One ensures that the cell requires more DHFR, which requirement is met by replication of teh selection gene, by selecting only for cells that can grow in teh presence of ever-greater MTX concentration. So long as the gene encoding a desired heterologous protein has cointegrated with the selection gene replication of this gene gives rise to replication of the gene encoding the desired protein. The result is that increased copies of the gene, i.e. an amplified gene, encoding the desired heterologous protein express more of the desired protein.

Suitable eukaryotic host cells for expressing the isolated polypeptide include monkey kidney CV1 line transformed by SV40 (COS-7, ATCC CRL 1651); human embryonic kidney line (293 or 293 cells subcloned for growth in suspension culture, Graham, F.L. *et al., J. Gen Virol.* 36: 59 (1977)); baby hamster kidney cells (BHK, ATCC CCL 10); chinese hamster ovary-cells-DHFR (CHO, Urlaub and Chasin, PNAS (USA) 77: 4216, [1980]); mouse sertoli cells (TM4, Mather, J.P., Biol. Reprod. 23: 243-251 [1980]); monkey kidney cells (CV1 ATCC CCL 70); african green monkey kidney cells (VERO-76, ATCC CRL-1587); human cervical carcinoma cells (HELA, ATCC CCL 2); canine kidney cells (MDCK, ATCC CCL 34); buffalo rat liver cells (BRL 3A, ATCC CRL 1442); human lung cells (W138, ATCC CCL 75); human liver cells (Hep G2, HB 8065); mouse mammary tumor (MMT 060562, ATCC CCL51); and, TRI cells (Mather, J.P. *et al., Annals N.Y. Acad. Sci.* 383: 44-68 [1982]).

Construction of suitable vectors containing the desired coding and control sequences employ standard ligation techniques. Isolated plasmids or DNA fragments are cleaved, tailored, and religated in the form desired to form the plasmids required.

For analysis to confirm correct sequences in plasmids constructed, the ligation mixtures are used to transform *E. coli* K12 strain 294 (ATCC 31446) and successful transformants selected by ampicillin or tetracycline resistance where appropriate. Plasmids from the transformants are prepared, analyzed by restriction and/or sequenced by the method of Messing *et al., Nucleic Acids Res.* 9: 309 (1981) or by the method of Maxam *et al., Methods in Enzymology* 65: 499 (1980).

Host cells are transformed with the expression vectors of this invention and cultured in conventional nutrient media modified as appropriate for inducing promoters, selecting transformants or amplifying the genes encoding the desired sequences. The culture

conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to the ordinarily skilled artisan.

The host cells referred to in this disclosure encompass cells in *in vitro* cultur as well as cells which are within a host animal.

"Transformation" means introducing DNA into an organism so that the DNA is replicable, either as an extrachromosomal element or by chromosomal integration. Unless indicated otherwise, the method used herein for transformation of the host cells is the method of Graham, F. and van der Eb, A., *Virology* 52: 456-457 (1973). However, other methods for introducing DNA into cells such as by nuclear injection or by protoplast fusion may also be used. If prokaryotic cells or cells which contain substantial cell wall constructions are used, the preferred method of transfection is calcium treatment using calcium chloride as described by Cohen, F.N. *et al., Proc. Natl. Acad. Sci. (USA),* 69: 2110 (1972).

"Transfection" refers to the introduction of DNA into a host cell whether or not any coding sequences are ultimately expressed. Numerous methods of transfection are known to the ordinarily skilled artisan, for example, $CaPO_4$ and electroporation. Transformation of the host cell is the indicia of successful transfection.

The novel polypeptide of this invention is recovered and purified from recombinant cell cultures by known methods, including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, immunoaffinity chromatography, hydroxyapatite chromatography and lectin chromatography. See, e.g., the purification methods described in EP 187,041. Moreover, reverse-phase HPLC and chromatography using ligands for the isolated polypeptide are useful for purification. It is presently preferred to utilize gel permeation chromatography and anion exchange chromatography, and more preferred to use cation exchange and hydrophobic interaction chromatography (HIC) according to standard protocols.

Optionally, the isolated polypeptide is recovered and purified by passage over a column of isolated polypeptide-antibody covalently coupled to aldehyde silica by a standard procedure (Roy *et al., Journal of Chromatography* 303:225-228 (1984)), washing of the column with a saline solution, and analyzing the eluant by standard methods such as quantitative amino acid analysis. Procedures utilizing monoclonal antibodies coupled to glycerol-coated controlled pore glass are desirable for the practice of this invention. Optionally, low concentrations (approximately 1-5 mM) of calcium ion may be present during purification. The isolated polypeptide may preferably be purified in the presence of a protease inhibitor such as PMSF.

The isolated polypeptide is placed into pharmaceutically acceptable, sterile, isotonic formulations together with required cofactors, and optionally are administered by standard means well known in the field. The formulation is preferably liquid, and is ordinarily a

physiologic salt solution containing non-phosphate buffer at pH 6.8-7.6, or may be lyophilized powder.

The isolated polypeptide compositions to be used in therapy will be formulated and dosages established in a fashion consistent with good medical practice taking into account the disorder to be treated, the condition of the individual patient, the site of delivery of the isolated polypeptide, the method of administration and other factors known to practitioners.

The isolated polypeptide is prepared for administration by mixing the isolated polypeptide at the desired degree of purity with adjuvants or physiologically acceptable carriers i.e. carriers which are nontoxic to recipients at the dosages and concentrations employed. Adjuvants and carriers are substances that in themselves share no immune epitopes with the target antigen, but which stimulate the immune response to the target antigen. Ordinarily, this will entail combining the isolated polypeptide with buffers, low molecular weight (less that about 10 residues) polypeptides, proteins, amino acids, carbohydrates including glucose or dextrans, chelating agents such as EDTA, and other excipients. Freunds adjuvant (a mineral oil emulsion) commonly has been used for this purpose, as have a variety of toxic microbial substances such as mycobacterial extracts and cytokines such as tumor necrosis factor and interferon gamma in U.S. patent 4,963,354. Although antigen is desirably administered with an adjuvant, in situations where the initial inoculation is delivered with an adjuvant, boosters with antigen may not require adjuvant. Carriers often act as adjuvants, but are generally distinguished from adjuvants in that carriers comprise water insoluble macromolecular particulate structures which aggregate the antigen, Typical carriers include aluminum hydroxide, latex particles, bentonite and liposomes.

It is envisioned that injections (intramuscular or subcutaneous) will be the primary route for therapeutic administration of the vaccines of this invention, intravenous delivery, or delivery through catheter or other surgical tubing is also used. Alternative routes include tablets and the like, commercially available nebulizers for liquid formulations, and inhalation of lyophilized or aerosolized receptors. Liquid formulations may be utilized after reconstitution from powder formulations.

The novel polypeptide may also be administered via microspheres, liposomes, other microparticulate delivery systems or sustained release formulations placed in certain tissues including blood. Suitable examples of sustained release carriers include semipermeable polymer matrices in the form of shaped articles, e.g. suppositories, or microcapsules. Implantable or microcapsular sustained release matrices include polylactides (U.S. Patent 3,773,919, EP 58,481) copolymers of L-glutamic acid and gamma ethyl-L-glutamate (U. Sidman *et al., Biopolymers* 22(1): 547-556, (1985)), poly (2-hydroxyethyl-methacrylate) or ethylene vinyl acetate (R. Langer *et al., J. Biomed. Mater. Res.* 15: 167-277 (1981) and R. Langer, *Chem. Tech.* 12: 98-105 (1982)). Liposomes containing the isolated polypeptide are prepared by well-known methods: DE 3,218,121A; Epstein *et al., Proc. Natl. Acad. Sci. USA*,

82:3688-3692 (1985); Hwang *et al., Proc. Natl. Acad. Sci. USA*, 77:4030-4034 (1980); EP 52322A; EP 36676A; EP 88046A; EP 143949A; EP 142541A; Japanese patent application 83-11808; U.S. Patents 4,485,045 and 4,544,545; and UP 102,342A. Ordinarily the liposomes are of the small (about 200-800 Angstroms) unilamelar type in which the lipid content is greater than about 30 mol. % cholesterol, the selected proportion being adjusted for the optimal rate of the polypeptide leakage.

The dose of the isolated polypeptide administered will be dependent upon the properties of the isolated polypeptide employed, e.g. its binding activity and in vivo plasma half-life, the concentration of the isolated polypeptide in the formulation, the administration route, the site and rate of dosage, the clinical tolerance of the patient involved, the pathological condition afflicting the patient and the like, as is well within the skill of the physician. Generally, doses of from about $0.5 \times 10^{*}$ to $5 \times 10^{*}$ molar of isolated polypeptide per patient per administration are preferred. Different dosages are utilized during a series of sequential inoculations; the practitioner may administer an initial inoculation and then boost with relatively smaller doses of isolated polypeptide vaccine.

The isolated polypeptide vaccines of this invention may be administered in a variety of ways and to different classes of recipients. The vaccines are used to vaccinate individuals who may or may not be at risk of exposure to HIV, and additionally, the vaccines are desirably administered to seropositive individuals and to individuals who have been previously exposed to HIV (see e.g. Salk, *Nature* 327:473-476 (1987); and Salk *et al., Science* 195:834-847 (1977)).

The isolated polypeptide may be administered in combination with other antigens in a single inoculation "cocktail". The isolated polypeptide vaccines may also be administered as one of a series of inoculations administered over time. Such a series may include inoculation with the same or different preparations of HIV antigens or other vaccines.

The adequacy of the vaccination parameters chosen, e.g. dose, schedule, adjuvant choice and the like, is determined by taking aliquots of serum from the patient and assaying antibody titers during the course of the immunization program. Alternatively, the presence of T cells may by monitored by conventional methods as described in Example 1 below. In addition, the clinical condition of the patient will be monitored for the desired effect, e.g. anti-infective effect. If inadequate vaccination is achieved then the patient can be boosted with further isolated polypeptide vaccinations and the vaccination parameters can be modified in a fashion expected to potentiate the immune response, e.g. increase the amount of antigen and/or adjuvant, complex the antigen with a carrier or conjugate it to an immunogenic protein, or vary the route of administration.

For use of the isolated polypeptide as a vaccine, it is currently preferred that at least three separate inoculations with isolated polypeptide be administered, with a second inoculation being administered more than two, preferably three to eight, and more preferably

approximately four weeks following the first inoculation. It is preferred that a third inoculation be administered several months later than the second "boost" inoculation, preferably at least more than five months following the first inoculation, more preferably six months to two years following the first inoculation, and even more preferably eight months to one year following the first inoculation. Periodic inoculations beyond the third are also desirable to enhance the patient's "immune memory". See Anderson *et al.*, *J. Infectious Diseases* 160(6):960-969 (Dec. 1989). Generally, infrequent immunizations with isolated polypeptide spaced at relatively long intervals is more preferred than frequent immunizations in eliciting maximum antibody responses, and in eliciting a protective effect.

The polypeptides of this invention may optionally be administered along with other pharmacologic agents used to treat AIDS or ARC or other HIV-related diseases and infections, such as AZT, CD4, antibiotics, immunomodulators such as interferon, anti-inflammatory agents, and anti-tumor agents.

Antibodies

This invention is also directed to monoclonal antibodies. In accordance with this invention, monoclonal antibodies specifically binding an epitope of an isolated polypeptide or antigenically active fragments thereof are isolated from continuous hybrid cell lines formed by the fusion of antigen-primed immune lymphocytes with myeloma cells. The antibodies of the subject invention are obtained through routine screening. An assay is used for screening monoclonal antibodies for their cytotoxic potential as ricin A chain containing immunotoxins. The assay involves treating cells with dilutions of the test antibody followed by a Fab fragment of a secondary antibody coupled to ricin A chain ('indirect assay'). The cytotoxicity of the indirect assay is compared to that of the direct assay where the monoclonal antibody is coupled to ricin A chain. The indirect assay accurately predicts the potency of a given monoclonal antibody as an immunotoxin and is thus useful in screening monoclonal antibodies for use as immunotoxins - see also Vitetta *et al.*, *Science* 238:1098-1104 (1987), and Weltman *et al.*, *Cancer Res.* 47:5552 (1987).

Monoclonal antibodies are highly specific, being directed against a single antigenic site. Furthermore, in contrast to conventional antibody (polyclonal) preparations which typically include different antibodies directed against different determinants (epitopes), each monoclonal antibody is directed against a single determinant on the antigen. Monoclonal antibodies are useful to improve the selectivity and specificity of diagnostic and analytical assay methods using antigen- antibody binding. A second advantage of monoclonal antibodies is that they are synthesized by the hybridoma culture, uncontaminated by other immunoglobulins. Monoclonal antibodies may be prepared from supernatants of cultured hybridoma cells or from ascites induced by intra-peritoneal inoculation of hybridoma cells into mice.

The hybridoma technique described originally by Kohler and Milstein, *Eur. J. Immunol.*, 6:511 (1976) has been widely applied to produce hybrid cell lines that secret high levels of monoclonal antibodies against many specific antigens.

In particular embodiments of this invention, an antibody is obtained by immunizing mice such as Balb/c or, preferably C57 BL/6, against an isolated polypeptide and screening for a clonal antibody that, when preincubated with the isolated polypeptide, prevents its binding to isolated polypeptide. Monoclonal antibodies may desirably have differences in affinity, immunoglobulin class, species of origin, or epitope; they may be antibodies which are expressed in recombinant cell culture or that are predetermined amino acid sequence variants of known antibodies, including chimeras of antibodies having a variable region directed against an isolated polypeptide, and a human constant region.

The route and schedule of immunization of the host animal or cultured antibody-producing cells therefrom are generally in keeping with established and conventional techniques for antibody stimulation and production. Applicants typically have employed mice as the test model although it is contemplated that any mammalian subject including human subjects or antibody producing cells therefrom can be manipulated according to the processes of this invention to serve as the basis for production of mammalian, including human, hybrid cell lines.

After immunization, immune lymphoid cells are fused with myeloma cells to generate a hybrid cell line which can be cultivated and subcultivated indefinitely, to produce large quantities of monoclonal antibodies. For purposes of this invention, the immune lymphoid cells selected for fusion are lymphocytes and their normal differentiated progeny, taken either from lymph node tissue or spleen tissue from immunized animals. Applicants prefer to employ immune spleen cells, since they offer a more concentrated and convenient source of antibody producing cells with respect to the mouse system. The myeloma cells provide the basis for continuous propagation of the fused hybrid. Myeloma cells are tumor cells derived from plasma cells.

It is possible to fuse cells of one species with another. However, it is preferred that the source of immunized antibody producing cells and myeloma be from the same species.

The hybrid cell lines can be maintained in culture *in vitro* in cell culture media. The cell lines of this invention can be selected and/or maintained in a composition comprising the continuous cell line in hypoxanthine-aminopterin thymidine (HAT) medium. In fact, once the hybridoma cell line is established, it can be maintained on a variety of nutritionally adequate media. Moreover, the hybrid cell lines can be stored and preserved in any number of conventional ways, including freezing and storage under liquid nitrogen. Frozen cell lines can be revived and cultured indefinitely with resumed synthesis and secretion of monoclonal antibody. The secreted antibody is recovered from tissue culture supernatant by conventional methods such as precipitation, Ion exchange chromatography, affinity chromatography, or

the like. The antibodies described herein are also recovered from hybridoma cell cultures by conventional methods for purification of IgG or IgM as the case may be that heretofore have been used to purify these immunoglobulins from pooled plasma, e.g. ethanol or polyethylene glycol precipitation procedures. The purified antibodies are sterile filtered, and optionally are conjugated to a detectable marker such as an enzyme or spin label for use in diagnostic assays of isolated polypeptide in test samples.

While the invention covers using mouse monoclonal antibodies, the invention is not so limited; in fact, human antibodies may be used and may prove to be preferable. Such antibodies can be obtained by using human hybridomas (Cote *et al.*, *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, p. 77 (1985)). In fact, according to the invention, techniques developed for the production of chimeric antibodies (Morrison *et al.*, *Proc. Natl. Acad. Sci.*, 81:6851 (1984); Neuberger *et al.*, *Nature* 312:604 (1984); Takeda *et al.*, *Nature* 314:452 (1985)) by splicing the genes from a mouse antibody molecule of appropriate antigen specificity together with genes from a human antibody molecule of appropriate biological activity (such as ability to activate human complement and mediate ADCC) can be used; such antibodies are within the scope of this invention.

As another alternative to the cell fusion technique, EBV-immortalized B cells are used to produce the monoclonal antibodies of the subject Invention. Other methods for producing monoclonal antibodies such as recombinant DNA, are also contemplated.

Immunotoxins

This invention is also directed to immunochemical derivatives of the antibodies of this invention such as immunotoxins (conjugates of the antibody and a cytotoxic moiety). The antibodies are also used to induce lysis through the natural complement process, and to interact with antibody dependent cytotoxic cells normally present.

Purified, sterile filtered antibodies are optionally conjugated to a cytotoxin such as ricin for use in AIDS therapy. EPO Publication 0 279 688 published 24 August 1988 illustrates methods for making and using immunotoxins for the treatment of HIV infection.

Immunotoxins of this invention, capable of specifically binding regions of HIV env, are used to kill cells that are already infected and are actively producing new virus. Killing is accomplished by the binding of the immunotoxin to viral coat protein which is expressed on infected cells. The immunotoxin is then internalized and kills the cell. Infected cells that have incorporated viral genome into their DNA but are not synthesizing viral protein (i.e., cells in which the virus is latent) may not be susceptible to killing by immunotoxin until they begin to synthesize virus. The antibodies of this invention which span the clip site and/or the other antibodies described herein may be used alone or in any combination with for delivering toxins to infected cells. In addition, a toxin-antibody conjugate can bind to circulating viruses or viral coat protein which will then effect killing of cells that internalize virus or coat protein.

The subject invention provides a highly selective method of destroying HIV infected cells, utilizing the antibodies described herein.

While not wishing to be constrained to any particular theory of operation of the invention, it is believed that the expression of the target antigen on the infected cell surface is transient. The antibodies must be capable of reaching the site on the cell surface where the antigen resides and interacting with it. After the antibody complexes with the antigen, endocytosis takes place carrying the toxin into the cell.

The immunotoxins of this invention are particularly helpful in killing monocytes/macrophages infected with the HIV virus. In contrast to the transient production of virus from T cells, macrophages produce high levels of virus for long periods of time. Current therapy is ineffective in inhibiting the production of new viruses in these cells.

Not all monoclonal antibodies specific for an isolated polypeptide make highly cytotoxic immunotoxins, however assays are routinely and commonly used in the field to predict the ability of an antibody to function as part of a immunotoxin. Preferably the antibodies used cross react with several (or all) strains of HIV.

The cytotoxic moiety of the immunotoxin may be a cytotoxic drug or an enzymatically active toxin of bacterial, fungal, plant or animal origin, or an enzymatically active fragment of such a toxin. Enzymatically active toxins and fragments thereof used are diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, *Aleurites fordii* proteins, dianthin proteins, *Phytolaca americana* proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, sapaonaria officinalis inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin and the tricothecenes. In another embodiment, the antibodies are conjugated to small molecule anticancer drugs such as cis-platin or 5FU. Conjugates of the monoclonal antibody and such cytotoxic moieties are made using a variety of bifunctional protein coupling agents. Examples of such reagents are SPDP, IT, bifunctional derivatives of imidoesters such as dimethyl adipimidate HCl, active esters such as disuccinimidyl suberate, aldehydes such as glutaraldehyde, bis-azido compounds such as bis (p-azidobenzoyl) hexanediamine, bis-diazonium derivatives such as bis- (p-diazoniumbenzoyl)- -ethylenediamine, diisocyanates such as tolylene 2,6-diisocyanate and bis-active fluorine compounds such as 1,5-difluoro- 2,4-dinitrobenzene. The lysing portion of a toxin may be joined to the Fab fragment of the antibodies.

Immunotoxins can be made in a variety of ways, as discussed herein. Commonly known crosslinking reagents can be used to yield stable conjugates.

Advantageously, monoclonal antibodies specifically binding the domain of the protein which is exposed on the infected cell surface, are conjugated to ricin A chain. Most advantageously the ricin A chain is deglycosylated and produced through recombinant means.

An advantageous method of making the ricin immunotoxin is described in Vitetta *et al.*, *Science* 238:1098 (1987).

When used to kill infected human cells *in vitro* for diagnostic purposes, the conjugates will typically be added to the cell culture medium at a concentration of at least about 10 nM. The formulation and mode of administration for *in vitro* use are not critical. Aqueous formulations that are compatible with the culture or perfusion medium will normally be used. Cytotoxicity may be read by conventional techniques.

Cytotoxic radiopharmaceuticals for treating infected cells may be made by conjugating radioactive isotopes (e.g. I, Y, Pr) to the antibodies. Advantageously alpha particle-emitting isotopes are used. The term 'cytotoxic moiety" as used herein is intended to include such isotopes.

In a preferred embodiment, ricin A chain is deglycosylated or produced without oligosaccharides, to decrease its clearance by irrelevant clearance mechanisms (e.g., the liver). In another embodiment, whole ricin (A chain plus B chain) is conjugated to antibody if the galactose binding property of B-chain can be blocked ("blocked ricin").

In a further embodiment toxin-conjugates are made with Fab or $F(ab')_2$ fragments. Because of their relatively small size these fragments can better penetrate tissue to reach infected cells.

In another embodiment, fusogenic liposomes are filled with a cytotoxic drug and the liposomes are coated with antibodies specifically binding HIV env.

Antibody Dependent Cellular Cytotoxicity

The present invention also involves a method based on the use of antibodies which are (a) directed against an isolated polypeptide, and (b) belong to a subclass or isotype that is capable of mediating the lysis of HIV virus infected cells to which the antibody molecule binds. More specifically, these antibodies should belong to a subclass or isotype that, upon complexing with cell surface proteins, activates serum complement and/or mediates antibody dependent cellular cytotoxicity (ADCC) by activating effector cells such as natural killer cells or macrophages.

The present invention is also directed to the use of these antibodies, in their native form, for AIDS therapy. For example, IgG2a and IgG3 mouse antibodies which bind HIV-associated cell surface antigens can be used *in vitro* for AIDS therapy. In fact, since HIV env is present on infected monocytes and T-lymphocytes, the antibodies disclosed herein and their therapeutic use have general applicability.

Biological activity of antibodies is known to be determined, to a large extent, by the Fc region of the antibody molecule (Uananue and Benacerraf, *Textbook of Immunology*, 2nd Edition, Williams & Wilkins, p. 218 (1984)). This includes their ability to activate complement and to mediate antibody-dependent cellular cytotoxicity (ADCC) as effected by leukocytes. Antibodies of different classes and subclasses differ in this respect, and, according to the

present invention, antibodies of those classes having the desired biological activity are selected. For example, mouse immunoglobulins of the IgG3 and IgG2a class are capable of activating serum complement upon binding to the target cells which express the cognate antigen.

In general, antibodies of the IgG2a and IgG3 subclass and occasionally IgG1 can mediate ADCC, and antibodies of the IgG3, IgG2a, and IgM subclasses bind and activate serum complement. Complement activation generally requires the binding of at least two IgG molecules in close proximity on the target cell. However, the binding of only one IgM molecule activates serum complement.

The ability of any particular antibody to mediate lysis of the target cell by complement activation and/or ADCC can be assayed. The cells of interest are grown and labeled *in vitro*; the antibody is added to the cell culture in combination with either serum complement or immune cells which may be activated by the antigen antibody complexes. Cytolysis of th target cells is detected by the release of label from the lysed cells. In fact, antibodies can be screened using the patient's own serum as a source of complement and/or immune cells. The antibody that is capable of activating complement or mediating ADCC in the *in vitro* test can then be used therapeutically in that particular patient.

Antibodies of virtually any origin can be used for this purpose provided they bind an isolated polypeptide epitope and can activate complement or mediate ADCC. Monoclonal antibodies offer the advantage of a continuous, ample supply.

Therapeutic and Other Uses of the Antibodies

When used *in vivo* for therapy, the antibodies of the subject invention are administered to the patient in therapeutically effective amounts (i.e. amounts that restore T cell counts). They will normally be administered parenterally. The dose and dosage regimen will depend upon the degree of the infection, the characteristics of the particular immunotoxin (when used), e.g., its therapeutic index, the patient, and the patient's history. Advantageously the immunotoxin is administered continuously over a period of 1-2 weeks, intravenously to treat cells in the vasculature and subcutaneously and intraperitoneally to treat regional lymph nodes. Optionally, the administration is made during the course of adjunct therapy such as combined cycles of tumor necrosis factor and interferon or other immunomodulatory agent.

For parenteral administration the antibodies will be formulated in a unit dosage injectable form (solution, suspension, emulsion) in association with a pharmaceutically acceptable parenteral vehicle. Such vehicles are inherently nontoxic, and non-therapeutic. Examples of such vehicles are water, saline, Ringer's solution, dextrose solution, and 5% human serum albumin. Nonaqueous vehicles such as fixed oils and ethyl oleate can also be used. Liposomes may be used as carriers. The vehicle may contain minor amounts of additives such as substances that enhance isotonicity and chemical stability, e.g., buffers and

preservatives. The antibodies will typically be formulated in such vehicles at concentrations of about 1 mg/ml to 10 mg/ml.

Use of IgM antibodies is not currently preferred, since the antigen is highly specific for the target cells and rarely occurs on normal cells. IgG molecules by being smaller may be more able than IgM molecules to localize to infected cells.

There is evidence that complement activation *in vivo* leads to a variety of biological effects, including the induction of an inflammatory response and the activation of macrophages (Uananue and Benecerraf, *Textbook of Immunology*, 2nd Edition, Williams & Wilkins, p. 218 (1984)). The increased vasodilation accompanying inflammation may increase the ability of various anti-AIDS agents to localize in infected cells. Therefore, antigen-antibody combinations of the type specified by this invention can be used therapeutically in many ways. Additionally, purified antigens (Hakomori, *Ann. Rev. Immunol.* 2:103 (1984)) or anti-idiotypic antibodies (Nepom *et al.*, *Proc. Natl. Acad. Sci.* 81:2864 (1985); Koprowski *et al.*, *Proc. Natl. Acad. Sci.* 81:216 (1984)) relating to such antigens could be used to induce an active immune response in human patients. Such a response includes the formation of antibodies capable of activating human complement and mediating ADCC and by such mechanisms cause infected cell destruction.

The antibodies of the subject invention are also useful in the diagnosis of HIV in test samples. They are employed as one axis of a sandwich assay for an isolated polypeptide of HIV env, together with a polyclonal or monoclonal antibody directed at another sterically-free epitope of HIV env. For use in some embodiments of sandwich assays the anti-isolated polypeptide antibody is bound to an insolubilizing support or is labelled with a detectable moiety following conventional procedures used with other monoclonal antibodies. In another embodiment a labelled antibody, e.g. labelled goat anti-murine IgG, capable of binding the anti-isolated polypeptide antibody is employed to detect the isolated polypeptide or HIV env binding using procedures previously known *per se*.

The antibody compositions used in therapy are formulated and dosages established in a fashion consistent with good medical practice taking into account the disorder to be treated, the condition of the individual patient, the site of delivery of the composition, the method of administration and other factors known to practitioners. The antibody compositions are prepared for administration according to the description of preparation of polypeptides for administration, *infra*.

In order to facilitate understanding of the following examples certain frequently occurring methods and/or terms will be described.

"Plasmids" are designated by a lower case p preceded and/or followed by capital letters and/or numbers. The starting plasmids herein are either commercially available, publicly available on an unrestricted basis, or can be constructed from available plasmids in accord

with published procedures. In addition, equivalent plasmids to those described are known in the art and will be apparent to the ordinarily skilled artisan.

In particular, it is preferred that these plasmids have some or all of the following characteristics: (1) possess a minimal number of host-organism sequences; (2) be stable in the desired host; (3) be capable of being present in a high copy number in the desired host; (4) possess a regulatable promoter; and (5) have at least one DNA sequence coding for a selectable trait present on a portion of the plasmid separate from that where the novel DNA sequence will be inserted. Alteration of plasmids to meet the above criteria are easily performed by those of ordinary skill in the art in light of the available literature and the teachings herein. It is to be understood that additional cloning vectors may now exist or will be discovered which have the above-identified properties and are therefore suitable for use in the present invention and these vectors are also contemplated as being within the scope of this invention.

"Digestion" of DNA refers to catalytic cleavage of the DNA with a restriction enzyme that acts only at certain sequences in the DNA. The various restriction enzymes used herein are commercially available and their reaction conditions, cofactors and other requirements were used as would be known to the ordinarily skilled artisan. For analytical purposes, typically 1 $\mu$g of plasmid or DNA fragment is used with about 2 units of enzyme in about 20 $\mu$l of buffer solution. For the purpose of isolating DNA fragments for plasmid construction, typically 5 to 50 $\mu$g of DNA are digested with 20 to 250 units of enzyme in a larger volume. Appropriate buffers and substrate amounts for particular restriction enzymes are specified by the manufacturer. Incubation times of about 1 hour at 37°C are ordinarily used, but may vary in accordance with the supplier's instructions. After digestion the reaction is electrophoresed directly on a polyacrylamide gel to isolate the desired fragment.

Size separation of the cleaved fragments is performed using 8 percent polyacrylamide gel described by Goeddel, D. *et al., Nucleic Acids Res.* 8: 4057 (1980).

"PCR" (polymerase chain reaction) refers to a technique whereby a piece of DNA is amplified. Oligonucleotide primers which correspond to the 3' and 5' ends (sense or antisense strand-check) of the segment of the DNA to be amplified are hybridized under appropriate conditions and the enzyme Taq polymerase, or equivalent enzyme, is used to synthesize copies of the DNA located between the primers.

"Dephosphorylation" refers to the removal of the terminal 5' phosphates by treatment with bacterial alkaline phosphatase (BAP). This procedure prevents the two restriction cleaved ends of a DNA fragment from "circularizing" or forming a closed loop that would impede insertion of another DNA fragment at the restriction site. Procedures and reagents for dephosphorylation are conventional. Maniatis, T. *et al., Molecular Cloning* pp. 133-134 (1982). Reactions using BAP are carried out in 50mM Tris at 68°C to suppress the activity

of any xonucleases which are present in the enzyme preparations. Reactions are run for 1 hour. Following the reaction the DNA fragment is gel purified.

"Oligonucleotides" refers to either a single stranded polydeoxynucleotide or two complementary polydeoxynucleotide strands which may be chemically synthesized. Such synthetic oligonucleotides have no 5' phosphate and thus will not ligate to another oligonucleotide without adding a phosphate with an ATP in the presence of a kinase. A synthetic oligonucleotide will ligate to a fragment that has not been dephosphorylated.

"Ligation" refers to the process of forming phosphodiester bonds between two double stranded nucleic acid fragments (Maniatis, T. *et al.., Id.*, p. 146). Unless otherwise provided, ligation is accomplished using known buffers and conditions with 10 units of T4 DNA ligase ("ligase") per 0.5 $\mu$g of approximately equimolar amounts of the DNA fragments to be ligated.

"Filling" or "blunting" refers to the procedures by which the single stranded end in the cohesive terminus of a restriction enzyme-cleaved nucleic acid is converted to a double strand. This eliminates the cohesive terminus and forms a blunt end. This process is a versatile tool for converting a restriction cut end that may be cohesive with the ends created by only one or a few other restriction enzymes into a terminus compatible with any blunt-cutting restriction endonuclease or other filled cohesive terminus. Typically, blunting is accomplished by incubating 2-15 $\mu$g of the target DNA in 10mM $MgCl_2$, 1mM dithiothreitol, 50mM NaCl, 10mM Tris (pH 7.5) buffer at about 37°C in the presence of 8 units of the Klenow fragment of DNA polymerase I and 250 $\mu$M of each of the four deoxynucleoside triphosphates. The incubation generally is terminated after 30 min. phenol and chloroform extraction and ethanol precipitation.

It is understood that the application of the teachings of the present invention to a specific problem or situation will be within the capabilities of one having ordinary skill in the art in light of the teachings contained herein. Examples of the products of the present invention and representative processes for their isolation, use, and manufacture appear below, but should not be construed to limit the invention.

## EXAMPLE

We have been able to produce large amounts of two different rgp120 fusion proteins in a mammalian cell system (Lasky *et al.*, 1986). This has allowed us to elucidate all nine of the disulfide bonds, the positions of the glycosylation sites that are utilized and the type of oligosaccharide moiety present at each site in rgp120 from the $III_B$ isolate of HIV-1 produced in CHO cells.

This example describes the structural characterization of the recombinant envelope glycoprotein (rgp120) of human immunodeficiency virus type 1 produced by expression in Chinese hamster ovary cells. Enzymatic cleavage of rgp120 and reversed-phase high performance liquid chromatography were used to confirm the primary structure of the protein, to assign intrachain disulfide bonds and to characterize potential sites for N-glycosylation.

All of the tryptic peptides identified were consistent with the primary structure predicted from the cDNA sequence. Tryptic mapping studies combined with treatment of isolated peptides with *S. aureus* V8 protease or with peptide: N-glycosidase F (PNGase F) followed by ndoproteinase Asp-N permitted the assignment of all nine intrachain disulfide bonds of rgp120. The 24 potential sites for N-glycosylation were characterized by determining the susceptibilities of the attached carbohydrate structures to PNGase F and to endo-β-N-acetylglucosaminidase H. Tryptic mapping of enzymatically deglycosylated rgp120 was used in conjunction with Edman degradation and fast atom bombardment-mass spectrometry of individually treated peptides to determine which of these sites are glycosylated and what types of structures are present. The results indicate that all 24 sites of gp120 are utilized, including 13 that contain complex-type oligosaccharides as the predominant structures, and 11 that contain primarily high mannose-type and/or hybrid-type oligosaccharide structures.

For convenience, complete bibliographic references are given at the end of this Example.

EXPERIMENTAL PROCEDURES

The abbreviations used throughout this example are: AAA, amino acid analysis; AIDS, acquired immunodeficiency syndrome; amu, atomic mass unit; CHO, Chinese hamster ovary; DTT, dithiothreitol; endo H, endo-β-N-acetylglucosaminidase H; FAB-MS, fast atom bombardment-mass spectrometry; gD1, herpes simplex type 1 glycoprotein D; gp, glycoprotein; HIV, human immunodeficiency virus; HPLC, high performance liquid chromatography; IAA, iodoacetic acid; PNGase F, peptide: N-glycosidase F; PTH, phenylthiohydantoin; RCM, reduced and S-carboxymethylated; rgp, recombinant glycoprotein; SIV, simian immunodeficiency virus; TFA, trifluoroacetic acid; TPCK, L-1-*p*-tosylamido-2-phenylethyl chloromethyl ketone.

*Materials*-- Recombinant gp120 proteins were produced in CHO cells and purified by immunoaffinity chromatography as previously described (Lasky *et al.*, 1986). DTT, IAA, and 2-acetamido-1-β-(L-aspartamido)-1,2-dideoxy-D-glucose (GlcNAc-Asn) were obtained from Sigma Chemical Company. HPLC/Spectro Grade trifluoroacetic acid (Pierce), Acetonitrile UV (American B&J), and Milli Q™ water (Millipore) were used for reversed-phase HPLC. The enzymes used were TPCK trypsin from Worthington Biomedical Corp., endoproteinase Asp-N ("sequencing grade") obtained from Boehringer Mannheim GmbH, *S. aureus* V8 protease from ICN ImmunoBiologicals, and PNGase F (N-Glycanase™) and endo H from Genzyme.

*Reduction and S-Carboxymethylation*-- Recombinant gp120 (2.0 mg of CL44 [SEQ. ID NO. 12]) was dialyzed against 0.36 M Tris buffer, pH 8.6, containing 8 M urea and 3 mM EDTA. DTT was added to a concentration of 10 mM and the sample was incubated for 4 hours at ambient temperature. The sample was then treated with 25 mM IAA in the dark for 30

minutes at ambient temperature. The reaction was quenched with excess DTT, the sample was dialyzed against 0.1 M ammonium bicarbonate, and then lyophilized.

*Treatment of RCM rgp120 with PNGase F*-- RCM rgp120 (0.5 mg) was reconstituted in 0.1 ml of 0.25 M sodium phosphate, pH 8.6, containing 10 mM EDTA and 0.02% $NaN_3$ to a concentration of 5 mg per mL. Tryptic peptides were reconstituted to the same molar concentration in 0.05 mM sodium phosphate, pH 7.0, containing 0.02% $NaN_3$. PNGase F was added to the sample in the ratio of 12.5 units per mg of protein and the sample was incubated overnight at 37°C. RCM rgp120 treated with PNGase F was dialyzed against 0.1 M ammonium bicarbonate.

*Treatment of RCM rgp120 with Endo H*-- RCM rgp120 (0.5 mg) was reconstituted in 0.1 ml of 0.05 M sodium phosphate, pH 6.0, containing 0.02% $NaN_3$. Endo H (2 units/ml) was added to the sample in the ratio of 0.1 unit per mg of protein and the sample was incubated overnight at 37°C. RCM rgp120 treated with endo H was dialyzed against 0.1 M ammonium bicarbonate.

*Treatment with TPCK-Trypsin*-- Samples of untreated, PNGase F-treated and endo H-treated RCM rgp120 (0.5 mg aliquots of CL44 [SEQ. ID NO. 12]) in 0.1 M ammonium bicarbonate were treated at ambient temperature with TPCK-trypsin by the addition of aliquots of enzyme (enzyme to substrate ratio of 1:100 w/w) at 0 and 6 hours of incubation. The digestion was stopped after 24 hours by freezing the samples. For disulfide determinations, a sample of rgp120 (0.5 mg of 9AA [SEQ. ID NO. 11]) was treated with TPCK-trypsin using the same conditions.

*Treatment of Tryptic Peptides with PNGase F Followed by Endoproteinase Asp-N*-- Peptides (ranging from 0.5 nmol to 3.7 nmol) purified by reversed-phase HPLC of a 9AA tryptic digest were reconstituted in 0.05 M sodium phosphate, pH 7.0, containing 0.02% $NaN_3$ (0.05 ml). PNGase F (5 units in 0.06 ml of 0.05 M sodium phosphate, pH 7.0, containing 0.02% NaN3) was added and the samples were incubated for 20 hours at 37°C. Endoproteinase Asp-N (2 microgram) was then added and the samples were incubated for 20 hours at 37°C.

*Treatment of Tryptic Peptides with S. aureus V8 Protease*-- Peptides (3.0 nmol) purified by reversed-phase HPLC of a 9AA tryptic digest were reconstituted in 0.05 M sodium phosphate, pH 7.0, containing 0.02% $NaN_3$ (0.04 ml). V8 protease (5 microgram) was added at 0 and 7 hours and the sample was incubated for 24 hours at 37°C.

*Treatment of CL44 Peptides with Endo H Followed by PNGase F*-- Peptides (typically 3 nmol) purified by reversed-phase HPLC were reconstituted in 0.05 M sodium phosphate, pH 6.0, containing 0.02% $NaN_3$ (0.1 ml). Endo H (0.05 unit in 0.025 ml of 0.05 M sodium phosphate, pH 6.0, containing 0.02% $NaN_3$) was added and the sample was incubated for 20 hours at 37°C. PNGase F (6.25 units) and 0.5 M sodium phosphate, pH 10.3, containing

0.02 M EDTA and 0.02% $NaN_3$ (0.125 ml) were then added and the sample was incubated for 20 hours at 37°C.

*Reversed-phase HPLC*-- Tryptic digests were fractionated by reversed-phase HPLC on a 5 micron Vydac C18 endcapped column (4.6 mm x 250 mm). After equilibration with 0.1% aqueous TFA, the elution of tryptic peptides was carried out at 1 ml per minute with a linear gradient from 0 to 45% acetonitrile containing 0.08% TFA in 90 minutes. The system used was a Waters gradient liquid chromatograph consisting of two 6000A pumps, a 720 controller, and a WISP 710B injector, and a Perkin-Elmer LC75 single wavelength UV detector set at 214 nm.

Peptides subjected to further manipulations were fractionated by reversed-phase HPLC on a Vydac C18 column (2.1 mm x 250 mm) equilibrated in 0.1% aqueous TFA at a flow rate of 0.2 ml per minute and a temperature of 40°C. These peptides were eluted with a linear gradient from 0 to 60% acetonitrile (containing 0.08% TFA) in 60 minutes. The system used was a Hewlett-Packard 1090M liquid chromatograph.

*Peptide Identification*-- Peptides collected from reversed-phase HPLC were identified by AAA and/or N-terminal sequence analysis. Samples for AAA were treated with constant boiling HCl at 110°C in vacuo for either 24 or 72 hours, depending upon extent of glycosylation. The extended hydrolysis degrades glucosamine, which would otherwise interfere with quantitation of Ile and Leu. Analysis was performed on a Beckman Model 6300 amino acid analyzer with ninhydrin detection.

N-terminal sequence analysis was performed on an Applied Biosystems Model 477A/120A. The acetonitrile concentration in the equilibration buffer of the PTH analysis system was decreased from 10 to 9% to resolve the PTH derivative of GlcNAc-Asn from DTT.

*FAB-MS*-- FAB mass spectra were acquired on a JEOL HX110HF/HX110HF tandem mass spectrometer operated in a normal two-sector mode. FAB-MS was performed with 6 keV xenon atoms (10 mA emission current). Data were acquired over a mass range of 380-4000 amu.

RESULTS

Lasky *et al.* (1986) expressed gp120 in CHO cells as a fusion protein using the signal peptide of the herpes simplex gD1. Two such fusion proteins were used in this study. The recombinant glycoprotein used in most of this study (CL44 [SEQ. ID NO. 12]) was expressed as a 498-amino-acid fusion protein containing the first 27 residues of gD1 fused to residues 31-501 of gp120 (Lasky *et al.*, 1986). This construction lacks the first cysteine residue of mature gp120. Disulfide assignments were carried out on another recombinant fusion protein (9AA [SEQ. ID NO. 11]) which contains the first 9 residues of gD1 fused to residues 4-501 of gp120. This restores the first cysteine residue, Cys 24. Carboxy-terminal analysis of CL44 [SEQ. ID NO. 12] using carboxypeptidase digestions indicated that glutamic acid residue

479 is the carboxy terminus of the fully processed molecule secreted by CHO cells (data not shown). The amino acid sequences of these two constructions are given in Figure 1.

*RCM CL44 Tryptic Map*-- Reversed-phase HPLC tryptic mapping was used to confirm the primary structure of the molecule, to assign intrachain disulfide bonds and to characterize potential sites for N-glycosylation. In experiments not intended to give information about disulfides, the protein was RCM prior to digestion with trypsin. This treatment unfolds the protein and disrupts disulfide bonds, thereby resulting in smaller tryptic fragments than would be obtained with the native molecule.

The reversed-phase HPLC tryptic map of RCM CL44 is shown in Figure 2. Tryptic peptides were separated by reversed-phase HPLC using an acetonitrile/water system with TFA as the ionic modifier. As will be discussed below, much of the peak heterogeneity derives from the extremely high (approximately 50% of total mass) carbohydrate content of the molecule. Peaks were collected and subjected to AAA for identification (Table I). In some cases, N-terminal sequence analysis was used for confirmation (these peaks are indicated in Table I). The peaks not assigned a label in Figure 2 were not identified.

Table 1: Amino Acid Compositions of Peptides from the Tryptic Map of RCM CL44.
Peptides are listed according to elution position in the tryptic map of RCM CL44. Theoretical values derived for peptides labelled in Figure 1 are in parenthesis. Non-integral theoretical values are the result of coeluting peptides.

| Peptide(s) | T35 | T19 T36 | T25 | T17 T18 T38 | H3 | H2 T5,6 T26[a] | T22 | T15 | T21 | T33 | T34 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| CyA | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| CMC | 0.2 | 0.0 | 0.0 | 0.0 | 0.0 | 0.9(1) | 0.9(1) | 0.1 | 0.1 | 0.0 | 0.0 |
| Asx | 0.4 | 0.2 | 1.2(1) | 0.4 | 0.0 | 7.7(9) | 1.6(1) | 2.9(3) | 1.1(1) | 1.8(2) | 0.1 |
| Thr | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 2.3(3) | 0.2 | 0.9(1) | 0.0 | 0.0 | 0.0 |
| Ser | 0.0 | 0.0 | 0.6(1) | 0.6(1) | 0.0 | 1.5(3) | 0.6(1) | 0.5(1) | 0.0 | 0.0 | 0.6(1) |
| Glx | 0.1 | 0.0 | 0.9(1) | 2.0(2) | 0.0 | 2.1(2) | 1.0(1) | 0.1 | 0.0 | 0.1 | 1.0(1) |
| Pro | 0.0 | 1.0(1) | 0.0 | 0.0 | 0.0 | 0.9(1) | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 |
| Gly | 0.1 | 2.1(2) | 0.1 | 0.3 | 0.1 | 2.5(2) | 0.0 | 0.1 | 0.9(1) | 0.1 | 0.1 |
| Ala | 0.1 | 0.0 | 0.1 | 0.0 | 0.0 | 1.0(1) | 1.4(1) | 1.9(2) | 0.1 | 0.0 | 0.0 |
| Cys | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 |
| Val | 0.2 | 0.7(2) | 0.1 | 1.2(2) | 0.1 | 0.1 | 0.0 | 0.1 | 0.0 | 0.1 | 0.1 |
| Met | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.7(1) | 0.0 | 0.0 | 0.9(1) | 0.0 | 0.0 |
| Ile | 0.3 | 0.0 | 0.9(1) | 2.2(2) | 0.0 | 0.2 | 1.0(1) | 0.1 | 1.0(1) | 0.0 | 0.0 |
| Leu | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0(1) | 0.0 | 0.0 | 0.0 | 0.0 | 1.0(1) |
| Tyr | 0.8(1) | 0.4 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 | 0.0 | 1.0(1) |
| Phe | 0.0 | 0.0 | 0.0 | 0.0 | 1.0(1) | 1.0(1) | 0.2 | 0.9(1) | 0.1 | 0.0 | 0.0 |
| His | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 1.0(1) | 0.1 | 0.1 | 0.0 | 0.0 |
| Lys | 1.2(1) | 0.8(1) | 0.9(1) | 0.5 | 0.0 | 2.0(2) | 0.2 | 1.2(1) | 0.0 | 0.0 | 1.0(1) |
| Trp | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0(1) | 0.0 |
| Arg | 0.1 | 1.1(1) | 0.1 | 2.6(3) | 1.0(1) | 3.6(2) | 1.0(1) | 0.1 | 1.0(1) | 1.1(1) | 0.0 |

[a] Confirmed by N-terminal sequence analysis.

Table 1 Continued

| Peptide(s) | T12c | T4b | T7<br>T24[a] | T23,24 | T12b | T32 | T4a<br>T32[a] | T37 | H1<br>T8,9<br>T20[a] | T16<br>T27[a] |
|---|---|---|---|---|---|---|---|---|---|---|
| CyA | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| CMC | 0.0 | 0.8(1) | 0.2 | 0.2 | 0.9(1) | 0.3 | 0.1 | 0.1 | 0.2(0.2) | 0.3(0.3) |
| Asx | 0.1 | 0.3 | 1.8(2) | 2.6(2) | 0.3 | 5.2(6) | 1.6(1.8) | 0.3 | 0.8(0.7) | 1.5(1.5) |
| Thr | 0.0 | 0.3 | 0.9(1) | 1.3(1) | 0.2 | 0.6 | 0.7(0.5) | 0.9(1) | 1.0(1.1) | 1.5(1.8) |
| Ser | 0.1 | 0.8(1) | 0.2 | 0.3 | 0.1 | 0.8(1) | 0.3(0.3) | 0.3 | 0.5(0.7) | 0.2(0.3) |
| Glx | 0.2 | 0.6 | 1.1(1) | 0.6 | 0.1 | 2.0(2) | 0.6(0.6) | 1.1(1) | 0.3(0.2) | 0.9(0.9) |
| Pro | 0.1 | 0.3 | 0.2 | 0.2 | 0.9(1) | 1.0(1) | 0.8(0.8) | 1.9(2) | 0.2 | 0.3(0.3) |
| Gly | 0.2 | 0.5 | 0.3 | 0.4 | 1.1(1) | 4.7(5) | 1.4(1.5) | 1.0(1) | 1.0(1.1) | 0.1 |
| Ala | 1.0(1) | 0.0 | 0.3 | 0.9(1) | 1.8(2) | 0.1 | 0.0 | 0.9(1) | 1.6(1.8) | 0.0 |
| Cys | 0.0 | 0.1 | 0.0 | 0.2 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Val | 0.3 | 1.2(1) | 0.2 | 0.5 | 0.1 | 0.4 | 0.1 | 0.9(1) | 0.9(0.9) | 0.6(0.6) |
| Met | 0.0 | 0.0 | 1.7(2) | 0.0 | 0.0 | 0.7(1) | 0.2(0.3) | 0.0 | 0.0 | 0.0 |
| Ile | 1.0(1) | 0.3 | 1.0(1) | 0.5 | 0.1 | 1.1(1) | 0.4(0.3) | 1.2(1) | 1.4(1.5) | 2.4(2.7) |
| Leu | 1.0(1) | 0.9(1) | 1.0(1) | 0.9(1) | 0.0 | 0.1 | 1.0(1.0) | 0.9(1) | 0.6(0.6) | 0.3(0.3) |
| Tyr | 0.0 | 0.0 | 0.1 | 0.2 | 0.0 | 0.1 | 0.0 | 0.1 | 0.3(0.3) | 0.0 |
| Phe | 0.0 | 0.0 | 0.0 | 0.5 | 1.1(1) | 1.2(1) | 0.3(0.3) | 0.2 | 1.0(1.1) | 0.9(0.9) |
| His | 0.0 | 0.3 | 0.1 | 0.2 | 0.1 | 0.2 | 0.1 | 0.0 | 0.0 | 0.0 |
| Lys | 1.0(1) | 0.9(1) | 1.9(2) | 1.7(2) | 0.0 | 0.1 | 0.0 | 1.0(1) | 1.3(1.4) | 0.8(0.9) |
| Trp | 0.0 | 0.0 | 0.0(1) | 0.0(1) | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| Arg | 0.2 | 0.0 | 0.0 | 0.2 | 0.1 | 1.8(2) | 0.5(0.6) | 0.2 | 0.2(0.2) | 0.6(0.6) |

[a] Confirmed by N-terminal sequence analysis.

Table 1 Continued

| Peptide(s) | T30 | T4 T10 T11 T31[a] | T12a | T29 | T14 T28[a] | H4,T2' | T2' | T3 |
|---|---|---|---|---|---|---|---|---|
| CyA | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 | 0.0 |
| CMC | 0.1 | 2.1(2.5) | 0.1 | 0.2 | 2.0(2.1) | 0.5(1) | 0.9(1) | 1.1(1) |
| Asx | 0.2 | 4.0(4.3) | 0.7 | 1.3(1) | 5.2(4.9) | 8.5(9) | 8.6(9) | 3.9(4) |
| Thr | 0.2 | 4.7(6.1) | 0.4 | 0.5 | 4.6(5.8) | 3.5(4) | 3.3(4) | 0.0 |
| Ser | 0.6(1) | 2.1(4.0) | 0.7(1) | 0.3 | 3.4(5.4) | 0.6 | 0.2 | 1.3(2) |
| Glx | 1.1(1) | 1.7(1.5) | 1.2(1) | 2.7(3) | 4.5(4.5) | 5.5(5) | 5.1(5) | 4.0(4) |
| Pro | 2.0(2) | 1.6(1.7) | 1.9(2) | 0.2 | 1.6(1.6) | 3.9(4) | 4.0(4) | 1.1(1) |
| Gly | 1.1(1) | 0.9(0.8) | 0.3 | 1.1(1) | 4.1(4.2) | 4.5(1) | 0.2 | 0.1 |
| Ala | 1.9(2) | 1.5(1.5) | 0.1 | 0.1 | 0.4(0.3) | 2.1(2) | 2.0(2) | 0.1 |
| Cys | 0.0 | 0.1 | 0.0 | 0.0 | 0.0 | 0.0 | 0.1 | 0.0 |
| Val | 0.2 | 1.1(1.0) | 1.1(1) | 1.0(1) | 1.8(2.3) | 8.0(8) | 8.0(8) | 2.1(2) |
| Met | 0.8(1) | 0.0 | 0.0 | 0.7(1) | 0.0 | 0.6(1) | 0.9(1) | 1.6(2) |
| Ile | 1.9(2) | 3.2(3.6) | 1.9(2) | 0.9(1) | 1.5(1.6) | 0.3 | 0.2 | 1.9(2) |
| Leu | 0.1 | 4.3(4.7) | 0.3 | 0.3 | 2.1(2.2) | 3.2(3) | 3.1(3) | 2.0(2) |
| Tyr | 1.0(1) | 2.2(2.3) | 0.8(1) | 0.1 | 0.6(0.5) | 1.4(1) | 0.9(1) | 0.0 |
| Phe | 0.1 | 1.7(1.6) | 1.0(1) | 1.0(1) | 3.1(2.8) | 1.7(1) | 1.0(1) | 0.1 |
| His | 0.0 | 0.0 | 1.0(1) | 0.1 | 0.7(0.8) | 2.2(2) | 2.0(2) | 1.0(1) |
| Lys | 0.1 | 1.8(1.8) | 0.1 | 0.8(1) | 0.1 | 1.7(2) | 1.1(1) | 2.0(2) |
| Trp | 0.0 | 0.0 | 0.0 | 0.0(1) | 0.0(1.0) | 0.0(2) | 0.0(2) | 0.0(1) |
| Arg | 1.0(1) | 0.8(0.8) | 0.1 | 0.1 | 1.0(1.1) | 0.0 | 0.1 | 0.0 |

[a] Confirmed by N-terminal sequence analysis.

All of the peptides identified were consistent with the primary structure predicted from the cDNA sequence. Of the 38 predicted peptides with 3 or more amino acids, 36 were identified in the tryptic map of RCM CL44. In addition, 4 predicted peptides consisting of 2 amino acids each were also identified (H3, H4, T23, and T35). The tripeptide composed of residues 139-141 (VQK) was not identified in the map and was not given a label in Figure 2. The only other peptide not identified was T13 (CNNK). Asparagine residue 200 of peptide T13 is a potential glycosylation site and the peptide lacks hydrophobic amino acids. Therefore, this glycopeptide is likely to be extremely hydrophilic and poorly resolved from the salt fraction on the reversed-phase column.

Tryptic cleavage did not occur between peptides T5 and T6 and between peptides T8 and T9. These are designated in Figure 2 as two T-numbers separated by a comma (T5,6 and T8,9). The absence of cleavage was confirmed by N-terminal sequence analysis of the peptides. In both of these cases, the asparagine residue to the C-terminal side of the cleavage site is a potential N-glycosylation site and it is likely that the carbohydrate moiety interferes with the action of trypsin. Incomplete tryptic cleavage was also observed between peptides H4 and T2' and between peptides T23 and T24 (H4,T2' and T23,24).

Several peptides arising from non-tryptic cleavages were observed in the tryptic map of RCM CL44. Two of the predicted tryptic peptides were further cleaved by "chymotrypsin-like" cleavages. Peptide T12 was completely cleaved after tyrosine residue 187 and phenylalanine residue 193 to yield peptides T12a, T12b, and T12c. Peptide T4 was partially hydrolyzed after leucine residue 95 to yield peptides T4a and T4b. Intact peptide T4 was also present.

One of the tryptic peptides, T22 (QAHCNISR) [SEQ. ID NO. 14] eluted at two different positions (32.4 minutes and 34.1 minutes) in the RCM CL44 tryptic map. Deglycosylation studies (discussed below) with PNGase F and endo H indicated that the different retention times of the two forms of peptide T22 are not due to carbohydrate differences. It is possible that this retention time heterogeneity results from partial conversion of the N-terminal glutamine residue to pyroglutamic acid (Sanger and Thompson, 1953).

*Disulfide Assignments in gp120*-- Mature gp120 contains 18 cysteine residues (shaded in Figure 1) and therefore could contain 9 intrachain disulfide bonds. The CL44 [SEQ. ID NO. 12] construction lacks Cys-24, the first cysteine residue of gp120 (Lasky *et al.*, 1986); therefore, a different construction (9AA [SEQ. ID NO. 11]), in which the first cysteine residue was restored, was expressed and purified to approximately the same degree as CL44 (L. Riddle, T. Gregory and D. Dowbenko, unpublished data). Ellman's reagent (Ellman, 1959) was used to demonstrate the absence of free sulfhydryl groups in 9AA [SEQ. ID NO. 11] (data not shown). Therefore, disulfide assignments were determined for the 9AA construction.

Tryptic mapping studies performed without S-carboxymethylation of cysteine residues allowed partial assignment of disulfides. The tryptic map of 9AA is shown in Figure 3. Peaks were identified by N-terminal sequence analysis (Table II). These identifications allowed unequivocal assignment of three of the nin disulfide bonds: between Cys-101 and Cys-127 (Peak A, Table II); between Cys-266 and Cys-301 (Peak B, Table II); and between Cys-24 and Cys-44 (Peak E, Table II).

Peptides containing the remaining cysteine residues were also identified (Table II). Peptide T28 contains three cysteine residues and coelutes with peptide T31, which contains one cysteine residue (Peak D, Table II). Peptide T11 contains two cysteine residues and coelutes with peptides T3 and T4, each of which contains a single cysteine residue (Peak F, Table II). Similarly, peptide T14 contains two cysteine residues and coelutes with peptides T12 and T13, each of which has a single cysteine residue (Peaks C and E, Table II). In each of these cases more than one disulfide bond was present in the group of tryptic peptides, thereby preventing unambiguous assignment. These tryptic peptides were further manipulated as described below to introduce selective cleavage between cysteine residues located on a single peptide.

Table II. *Identification of Cysteine-containing Peptides from the Tryptic Map of 9AA.*

Cys-containing peaks from the tryptic map of 9AA were identified by N-terminal sequence analysis. Cysteines in boxes joined by a solid line represent disulfide bond assignments. Cysteines in boxes joined by dotted lines represent disulfide bonds that could not be assigned unambiguously in this experiment. Partial cleavages are indicated by a parenthesis. Cysteines are labelled by an amino acid number and peptides are labelled with T-numbers corresponding to the nomenclature used in Figure 1.

| *Peak* | *Cys-Containing Peptides* |
|---|---|
| A | 101 (T5,6)<br>C TDLKNDTNTNSSSGR<br>(GEIK)N C SFNISTSIR<br>127 (T8,9) |
| B | 266 (T16)<br>TIIVQLNQSVEIN C TRPNNNTR<br>QAH C NISR<br>301 (T22) |
| C | (T12a,b) 188<br>VSFEPIPIHY C APAGF 198 (T13)<br>C NNK<br>(T14a)<br>TFNGTGP C TNVSTVQ C THGIR<br>209 217 |
| D | (T28) 348 355<br>QSSGGDPEIVTHSFN C GGEFFY C NSTQLFNSTWFNSTWSTE-<br>-GSNNTEGSDTITLP C R<br>388<br>(T31)<br>C SSNITGLLLTR<br>415 |
| E | 24 (T1)<br>EATTTLF C ASDAK<br>AYDTEVHNVWATHA C VPTDPNPQEVVLVNVTENFNMWK<br>44 (T2) |
| | (T12) 188<br>VSFEPIPIHY C APAGFAILK 198 (T13)<br>C NNK<br>(T14)<br>TFNGTGP C TNVSTVQ C THGIRPVVSTQLLLNGSLAEEEVVIR<br>209 217 |
| F | (T3) 89<br>NDMVEQMHEDIISLWDQSLKP C VK 96 (T4)<br>LTPL C VSLK<br>(T11) |

Each of the peptides has a potential N-linked glycosylation sit located between the cysteine residues. The peptides were treated with PNGase F, which removes asparagine-linked carbohydrate while converting the attachment asparagine residue to aspartic acid (Tarentino *et al.*, 1985). The resulting aspartic acid residue serves as a point for selective cleavage of the peptides with endoproteinase Asp-N (Drapeau, 1980). The peptides were separated by reversed-phase HPLC and identified by N-terminal sequence analysis.

The HPLC chromatogram obtained after treatment of peptides T12, T13, and T14 (Peak C, Figure 3) with PNGase F followed by endoproteinase Asp-N is given in Figure 4a, and the sequences of relevant peptides are given in Table III. The results indicate that rgp120 has disulfide bonds between Cys-198 and Cys-209 and between Cys-188 and Cys-217 (Table III). Treatment of peptides T3, T4, and T11 (Peak F, Figure 3) with PNGase F followed by endoproteinase Asp-N allowed the recovery of fragments that demonstrated the presence of disulfide bonds between Cys-89 and Cys-175 and between Cys-96 and Cys-166 (Figure 4b and Table III)

Table III. *Assignment of Disulfides from Peptides isolated in Figure 4.*

The tryptic peptides that could not be assigned unambiguously in Table II were further manipulated as described in Figure 4. Peaks were identified by N-terminal sequence analysis.

| *Peak* | *Sequence* |
|---|---|
| 1 | 198<br>C N<br>DGTGPC T<br>209 |
| 2 | 188<br>EPIPIHYC APAGF<br>DVSTVQC THG(IR)<br>217 |
| 3 | 89<br>DQSLKPC VK<br>DTSVITQAC PK<br>175 |
| 4 | 96<br>LTPLC VSLK<br>DDTTSYTLTSC<br>166 |
| 5 | 348<br>IVTHSFNC GGE<br>C SSNITGLLLTR<br>415 |
| 6 | 355<br>FFYC NSTQLFNSTWFNSTWSTE<br>TITLPC R<br>388 |

The last two disulfide bonds were assigned by treating peptides T28 and T31 (Peak D, Figure 3) with V8 protease to cleave to the carboxy side of the glutamic acid and aspartic acid residues (Drapeau *et al.*, 1972) located between the cysteine residues of T28. The chromatogram obtained after V8 protease digestion of T28 and T31 is given in Figure 4c and the sequences of the relevant peptides are given in Table III. The results demonstrated the presence of disulfide bonds between Cys-348 and Cys-415 and between Cys-355 and Cys-388.

Thus, the combined results of the tryptic mapping analysis and the further selective degradations permitted the assignment of all nine intrachain disulfide bonds of rgp120. Parallel experiments performed on CL44 [SEQ. ID NO. 12] produced similar results for the 8 disulfide bonds remaining in that construction (not shown). The disulfide bond assignments of rgp120 are summarized in Figure 6.

*Glycosylation Sites of gp120*– Mature gp120 contains 24 potential sites for N-glycosylation, as recognized by the sequence: Asn-Xaa-Ser(Thr) (Kornfeld and Kornfeld, 1985). These sites are indicated by a dot above the corresponding asparagine residue in Figure 1A. In the present study, tryptic mapping of enzymatically deglycosylated CL44 [SEQ. ID NO. 12] was used in conjunction with Edman degradation and FAB-MS of individually treated peptides to determine which of the 24 potential N-glycosylation sites are glycosylated and which contain less fully processed (i.e. high mannose-type or hybrid-type) oligosaccharides.

The two enzymes used for deglycosylation were PNGase F and endo H. PNGase F releases all types of N-linked oligosaccharide structures by cleavage of the $\beta$-aspartylglucosylamine linkage (Tarentino *et al.*, 1985). Endo H releases only high mannose-type and hybrid-type oligosaccharide structures by cleaving between the two core N-acetylglucosamine residues (Tai *et al.*, 1977). Deglycosylation of a peptide can be monitored by the increase in retention time of the peak corresponding to the glycopeptide in the reversed-phase elution profile. Thus, it was possible to determine which peptides were glycosylated by treatment with PNGase F and, on the basis of susceptibility to endo H, to distinguish those with attached high mannose-type and/or hybrid-type oligosaccharides as the predominant structures.

The 24 potential glycosylation sites of CL44 [SEQ. ID NO. 12] are contained in 14 tryptic glycopeptides. Thirteen of these glycopeptides were identified in the tryptic map of RCM CL44 (Figure 2). As mentioned above, T13 (CNNK) [SEQ. ID NO. 15] was not identified. The tryptic maps of PNGase F-treated RCM CL44 and endo H-treated RCM CL44 are compared with the RCM CL44 tryptic map in Figure 5. The peaks corresponding to glycopeptides are labelled in each of the three tryptic maps.

As would be expected for a heavily glycosylated molecule, treatment of RCM CL44 with PNGas F (Figure 5b) simplified the tryptic map significantly. Typically, the peaks corresponding to potential glycopeptides in the RCM CL44 tryptic map (Figure 5a) were broad

and often appeared as multiplets. Deglycosylation resulted in sharp, single peaks for each peptide, indicating that the glycopeptide peak multiplicity and broadness was due to carbohydrate heterogeneity.

All of the 13 potential glycopeptides that had been identified in the tryptic map of RCM CL44 were shifted to later retention times in the tryptic map of PNGase F-treated material. This demonstrates that at least 13 of the 24 potential sites are glycosylated. Peptide T28 was not recovered after deglycosylation. This peptide contains a large number of non-polar amino acids and, after removal of the hydrophilic carbohydrate moieties, may bind irreversibly to the HPLC column. As described above, peptide T22 elutes at 2 positions in the RCM CL44 tryptic map presumably as a result of conversion of the N-terminal glutamine to pyroglutamic acid. The retention times of both of the T22 peaks were altered in the deglycosylated material produced by treatment with both PNGase F and endo H, confirming that the difference between these forms of peptide T22 in the RCM CL44 tryptic map was not due to carbohydrate heterogeneity.

The tryptic map of endo H-treated RCM CL44 (Figure 5c) indicated that 6 of the 13 tryptic glycopeptides were endo H-susceptible (peptides T14, T16, T22, T24, T28, and T31). In addition, a small amount of peptide T15 showed endo H susceptibility. For each of these glycopeptides, the elution time of the endo H-treated glycopeptide was earlier than that of the corresponding PNGase F-treated glycopeptide. This is due to the hydrophilic N-acetylglucosamine residue that remains attached to the asparagine residue following endo H treatment. Peptide T16 was not identified in the tryptic map of endo H-treated RCM CL44. This peptide contains 3 potential glycosylation sites and was poorly recovered under any circumstances.

Conclusions as to the type of glycosylation present on each of the tryptic glycopeptides based on susceptibility to PNGase F and endo H are summarized in Table IV. Seven of the 13 glycopeptides identified in the tryptic map of RCM CL44 contain only a single glycosylation site and thus could be characterized unambiguously with regard to enzyme susceptibility. Peptides T2' (Asn-58), T26 (Asn-326), and T32 (Asn-433) were deglycosylated only by PNGase F and, therefore, contain attached complex-type oligosaccharide structures. Peptides T22 (Asn-302), T24 (Asn-309), and T31 (Asn-418) were susceptible to both PNGase F and endo H and, therefore, carry high mannose-type and/or hybrid-type oligosaccharide structures. Peptide T15 is only partially susceptible to endo H; therefore, Asn-246 carries primarily complex-type oligosaccharides but must also have some attached high mannose-type and/or hybrid-type oligosaccharide structures.

Table IV. *Assignment of Glycosylation Type to RCM CL44 Tryptic Peptides by Susceptibility to PNGase F and Endo H.*

Susceptibility to PNGase F or endo H was determined by an increase in the retention time of a peptide in the tryptic map of RCM CL44 . PNGase F releases all types of N-linked oligosaccharide structures, whereas endo H releases only high mannose and hybrid oligosaccharide structures.

| *Tryptic Peptide* [a] | *Glycosylation Sites (Asn Residue #)* | *Susceptible To PNGase F* | *Susceptible To Endo H* | *Glycosylation Type* |
|---|---|---|---|---|
| T2' | 58 | Yes | No | Complex |
| T6 | 106,111 | Yes | No | Complex [b] |
| T9 | 126,130 | Yes | No | Complex [b] |
| T11 | 156,167 | Yes | No | Complex [b] |
| T14 | 204,211,232 | Yes | Yes | High Mannose, Hybrid, and/or Complex [c] |
| T15 | 246 | Yes | Trace | Complex (Trace High Mannose and/or Hybrid) |
| T16 | 259,265,271 | Yes | Yes | High Mannose, Hybrid, and/or Complex [c] |
| T22 | 302 | Yes | Yes | High Mannose and/or Hybrid |
| T24 | 309 | Yes | Yes | High Mannose and/or Hybrid |
| T26 | 326 | Yes | No | Complex |
| T28 | 356,362,367,376 | Yes | Yes | High Mannose, Hybrid, and/or Complex [c] |
| T31 | 418 | Yes | Yes | High Mannose and/or Hybrid |
| T32 | 433 | Yes | No | Complex |

[a] T13 not found.

[b] Either or both sites glycosylated.

[c] Endo H susceptible glycosylation at one or more site(s).

Peptides T6, T9, and T11 each contain 2 potential glycosylation sites. Each peptide was deglycosylated by PNGase F but not by endo H indicating the presence of mostly complex-type oligosaccharide structures. In order to determine whether one or both of the potential glycosylation sites in each peptide were actually glycosylated, the PNGase F-treated glycopeptides were subjected to either FAB-MS or Edman degradation. Treatment with PNGase F converts the attachment asparagine residue to aspartic acid during deglycosylation (Tarentino *et al.*, 1985). This conversion can be detected by FAB-MS as an increase of 1 amu in the mass of the peptide for each site deglycosylated (Carr and Roberts, 1986) or by Edman degradation by the appearance of the PTH derivative of aspartic acid at the appropriate cycles. FAB-MS of deglycosylated peptide T5,6 revealed an ion corresponding to the peptide mass plus 2 amu ($[MH]^+$ observed: m/z 1772.6; calculated: m/z 1772.7). FAB-MS of deglycosylated peptide T9 gave similar results ($[MH]^+$ observed: m/z 1301.8; calculated: m/z 1301.5). Edman degradation was performed instead of FAB-MS on deglycosylated peptide T11 because of its high molecular weight (>2000 amu). Aspartic acid was observed in cycles 8 (derived from Asn-156) and 19 (derived from Asn-167). These combined results indicate the presence of complex-type oligosaccharide structures attached to Asn residues 106, 111, 126, 130, 156, and 167.

The remaining 3 glycopeptides identified in the tryptic map of RCM CL44 contained multiple potential glycosylation sites and were endo H-susceptible. Peptides T14, T16, and T28 account for a total of 10 potential glycosylation sites. Characterization of each glycosylation site was achieved by Edman degradation of HPLC-purified peptides that had been subjected to treatment with endo H followed by PNGase F.

When endo H releases the high mannose-type and hybrid-type oligosaccharide structures, it leaves an N-acetylglucosamine residue attached to the asparagine residue of the peptide (Tarentino *et al.*, 1974). PNGase F will not remove this N-acetylglucosamine residue, but will release the remaining N-linked oligosaccharide structures by cleavage of the $\beta$-aspartylglucosylamine bond, resulting in conversion of the attachment asparagine residue to aspartic acid (Chu, 1986). Therefore, treatment with Endo H followed by PNGase F will yield asparagine at an unglycosylated site, GlcNAc-Asn at a glycosylation site that contained primarily high mannose-type and/or hybrid-type oligosaccharide structures, and aspartic acid at a glycosylation site that carried primarily complex-type oligosaccharide structures. Paxton *et al.* (1987) have shown that it is possible to detect the PTH derivative of GlcNAc-Asn after Edman degradation. Using this approach, it was possible to characterize the remainder of the glycosylation sites of CL44 [SEQ. ID NO. 12]. For example, treatment of glycopeptide T16, which contains 3 potential N-glycosylation sites, with endo H followed by PNGase F resulted in the appearance of the PTH derivative of GlcNAc-Asn at cycles 7 and 13 and the appearance of PTH-Asp at cycle 19 during Edman degradation. Thus, glycopeptide T16 carries primarily high mannose-type and/or hybrid-type oligosaccharides at Asn-259 and

Asn-265 and complex-type oligosaccharides at Asn-271. The results of these experiments are summarized in Table V and indicate that CL44 [SEQ. ID NO. 12] contains complex-type oligosaccharide structures at Asn residues 271, 367, and 376 and high mannose-type and/or hybrid-type oligosaccharide structures at Asn residues 204, 211, 232, 259, 265, 356, and 362.

Table V. *Assignment of Glycosylation Type to RCM CL44 Tryptic Glycopeptides Containing Multiple Potential Glycosylation Sites.*

Characterization of multiple potential glycosylation sites on RCM CL44 tryptic glycopeptides was achieved by Edman degradation of HPLC purified peptides subjected to treatment with endo H followed by PNGase F. Edman degradation of deglycosylated peptides shows either an Asn residue at an unglycosylated site, a GlcNAc-Asn at a glycosylation site to which had been attached high mannose or hybrid oligosaccharide structures, or an Asp residue at a glycosylation site which had carried complex type oligosaccharide structures.

| *Tryptic Peptide* | *Asn Residue #* | *Residue Observed* | *Glycosylation Type* |
|---|---|---|---|
| T14 | 204 | GlcNAc-Asn | High Mannose and/or Hybrid |
| | 211 | GlcNAc-Asn | High Mannose and/or Hybrid |
| | 232 | GlcNAc-Asn | High Mannose and/or Hybrid |
| T16 | 259 | GlcNAc-Asn | High Mannose and/or Hybrid |
| | 265 | GlcNAc-Asn | High Mannose and/or Hybrid |
| | 271 | Asp | Complex |
| T28 | 356 | GlcNAc-Asn | High Mannose and/or Hybrid |
| | 362 | GlcNAc-Asn | High Mannose and/or Hybrid |
| | 367 | Asp | Complex |
| | 376 | Asp | Complex |

Peptide T13, which contains the remaining glycosylation site, was not identified in any of the tryptic maps pr sented in this paper. However, FAB-MS data obtained from the void peak f a tryptic map of RCM CL44 treated with endo H followed by PNGase F revealed an ion corresponding to $MH^+$ for that peptide containing an attached N-acetylglucosamin residue (observed: m/z 740.1; calculated: m/z 740.4). The presence of peptide T13 in the void peak was further confirmed by AAA. Therefore, we conclude that Asn-200 is glycosylated and carries primarily high mannose-type and/or hybrid-type oligosaccharide structures.

The data presented here demonstrate that all 24 potential glycosylation sites of gp120 are utilized, that 13 sites contain primarily complex-type oligosaccharide structures while 11 sites contain primarily high mannose-type and/or hybrid-type oligosaccharide structures. The type of glycosylation at each site is summarized in Figure 6.

DISCUSSION

We have determined the disulfide bonding pattern and the attachment positions of oligosaccharide moieties of rgp120 from the $III_B$ isolate of HIV-1. A schematic representation of this information is presented in Figure 6 [SEQ. ID NO. 10]. The rgp120 molecules from which the structural data were obtained possess the functional properties attributed to gp120 produced by HIV-1 virions including high-affinity CD4 binding (Lasky *et al.*, 1987), and HIV-1 neutralizing antigenicity (Lasky *et al.*, 1986). We therefore conclude that the CHO-expressed gp120 is properly folded and that the disulfide-bonded domains reported here for the recombinant molecules are representative of those occurring in gp120 produced by HIV-1 virions.

*Functional Aspects of gp120 Structure*-- The gp120 molecule comprises five disulfide-bonded loop structures. The first and fourth are simple loops formed by single disulfide bonds while the second, third and fifth are more complex arrays of loops formed by nested disulfide bonds. The fourth disulfide-bonded domain (residues 266-301) has been shown to contain significant type-specific neutralizing epitopes (Matsushita *et al.*, 1988; Rusche *et al.*, 1988; Goudsmit *et al.*, 1988; Javaherian *et al.*, 1989) and the fifth disulfide-bonded domain (residues 348-415) has been shown to be important for CD4 binding (Lasky *et al.*, 1987; Kowalski *et al.*, 1987). No direct functional correlates have been described for the other three disulfide-bonded domains. The amino acid sequence of gp120 varies to a large extent between different viral isolates but the majority of the variability is localized in hypervariable regions which punctuate the otherwise relatively conserved sequences (Willey *et al.*, 1986; Modrow *et al.* 1987). Modrow *et al.* (1987) have identified five hypervariable regions which are characterized by sequence variation, insertions and deletions. Four of these hypervariable regions correspond to well delineated loops as indicated in Figure 6. With the exception of the third hypervariable loop (disulfide-bonded domain IV) the functional significance of these regions is unknown.

The positions of the cysteine residues and, presumably, the disulfide bonding pattern in gp120 are highly conserved between isolates. Among HIV-1 isolates, the only exception to this conservation is the Z3 isolate (Willey *et al.*, 1986) which has an additional pair of cysteine residues in the fourth hypervariable domain (residues 363-384). These residues most likely form a tenth disulfide bond in the gp120 from this isolate. The presence of this extra bond in such a hypervariable region probably has no more effect on the structure and function of the molecule than the other sequence variations that occur in that region.

As shown in Fig. 7 in HIV-2 [SEQ. ID NO. 13], and similarly in SIV (data not shown) the positions of the cysteine residues in disulfide-bonded domains I, II, IV and V are conserved (Human Retroviruses and AIDS (1989). G. Myeres, A. Rabson, S. Josephs, T. Smith, J. Berzofsky and F. Wong-Stahl, Editors. U.S. Government Printing Office, Los Alamos National Laboratory, Los Alamos, New Mexico, LA-UR, 89-743). In domain III there are two additional pairs of cysteine residues (three in SIV isolate MM142) which are presumed to be disulfide bonded within a finger-like domain III structure analogous to that illustrated in Figure 6. Another major difference between HIV-1, HIV-2 and SIV is that hypervariable region V2 is reduced to five amino acids in HIV-2 and SIV. The functional significance of the differences between HIV-1, HIV-2 and SIV is unknown at this time.

One of the most important functions of gp120 is its ability to bind to CD4 and thereby mediate the attachment of virions to susceptible cells (Klatzman *et al.*, 1984; Dalgleish *et al.*, 1984). The CD4-binding function has been localized by mutagenesis and structural studies (Lasky *et al.*, 1987; Kowalski *et al.*, 1987) to the region between residues 320 and 450, which includes the fifth disulfide-bonded domain. Lasky *et al.* (1987) showed that deletion of residues 396 to 407 and mutagenesis of Ala-402 to Asp abolished CD4 binding. They also mapped the epitope of a monoclonal antibody that blocks gp120-CD4 binding to residues 392-402. Kowalski *et al.* (1987) identified three regions as being involved with CD4 binding. Insertions between residues 333-334, 388-390 and 442-443 abolished CD4 binding. In addition, a deletion of residues 441-479 abolished CD4 binding while deletion of residues 362-369 within the fourth hypervariable region had no effect on binding. Cordonnier *et al.* (1989) have shown that mutagenesis of Trp-397 to Tyr or Phe decreases CD4 binding and changes to Ser, Gly, Val or Arg abolish binding. Nygren *et al.* (1988) have reported that a proteolytic fragment of gp120 from residue 322 to near the C-terminus retains the ability to bind to CD4. The results of these studies indicate that the CD4 binding capacity of gp120 is localized to the region between residues 320 and 450 and more specifically to the residues around 333-334, 442-443 and the sequence between 388 and 407.

In the course of efforts to map the epitope of monoclonal antibody 5C2-E5 which blocks gp120-CD4 binding, Lasky *et al.* (1987) treated rgp120 (CL44 [SEQ. ID NO 12]) with acetic acid to cleave the protein at aspartic acid residues (Ingram, 1963) and isolated the peptide fragment 383-426 from a column of immobilized anti-gp120 monoclonal antibody

5C2-E5. Digestion of reduced rgp120 yielded the same fragment. Consequently, it was concluded that a disulfide bond existed between Cys residues 388 and 415. In the analysis reported here we have failed to find this disulfide bond and, instead, have consistently found the disulfide bonds between Cys-355 and Cys-388, and between Cys-348 and Cys-415 as summarized in Figure 6. We believe that the true disulfide-bond assignment is as indicated in Figure 6 and that the acetic acid digestion produced some disulfide bond rearrangement (Ryle and Sanger, 1955) in the earlier work.

*The Oligosaccharides of gp120*– Approximately 50% of the apparent molecular mass of gp120 is carbohydrate. The structures of the oligosaccharide moieties released by hydrazinolysis of CL44 [SEQ. ID NO. 12] rgp120 have been exhaustively analyzed (Mizuochi *et al.*, 1988a; Mizuochi *et al.*, 1988b). These authors found that 33% of the N-linked oligosaccharides were of the high mannose-type, 4% were of the hybrid type, and 63% were of the complex type. Of the complex oligosaccharides 90% were fucosylated and 94% were sialylated. The complex structures were approximately 4% monoantennary, 61% biantennary, 19% triantennary and 16% tetraantennary. No O-linked oligosaccharides were found. Geyer *et al.* (1988) have analyzed the oligosaccharides of gp120 from the $III_B$ isolate of HIV-1-infected human cells. They found that high mannose-type oligosaccharides accounted for approximately 50% of the carbohydrate structures. The remaining structures were fucosylated, partially sialylated bi-, tri-, and tetraantennary complex-type oligosaccharides. No novel carbohydrate structures, or moieties that would be expected to act as heterophile antigens in man, have been isolated from gp120 from either source.

We have shown here that all 24 glycosylation sites are utilized, and that 13 of the 24 sites contain complex-type oligosaccharides as the predominant structures while 11 contain primarily hybrid and/or high mannose structures. The demonstration of endo H-susceptible structures at 11 of the 24 sites is consistent with the earlier results of Mizuochi *et al.* (1988a,1988b) who determined that nearly 40% of the total oligosaccharide structures released from rgp120 were hybrid and/or high mannose-type oligosaccharides.

The 24 potential N-linked glycosylation sites in the gp120 sequence are conserved to a large extent between different viral isolates (Willey *et al.*, 1986; Modrow *et al.*, 1987). Based on the gp120 sequence comparisons in these references, 13 of the sites on gp120 from the $III_B$ isolate of HIV-1 are absolutely conserved; these include 8 of the 11 sites that carry predominantly hybrid-type and/or high mannose-type oligosaccharides. Thus, the less fully processed (i.e. Endo H-susceptible) oligosaccharides of gp120 are found preferentially at the most conserved glycosylation sites. The remaining sites (8 complex and 3 hybrid/high mannose) are relatively conserved, even though many of them ccur in the hypervariable regi ns. The positions of these sites may shift or be deleted, but there is always ne or more new site(s) within 5 t 10 residues f the reference $III_B$ site. Studies by Willey *et al.* (1988)

demonstrated that mutagenesis of Asn-232 to Gln decreased the infectivity of virions containing the mutant gp120 molecules without affecting CD4 binding or syncytium formation. At this time, no particular functional significance can be attributed to the type of oligosaccharide structure at any of the sites.

The role of the carbohydrate moieties on gp120 in CD4 binding has been investigated by several authors (Lifson *et al.*, 1986; Matthews *et al.*, 1987; Fenouillet *et al.*, 1989). Those that employed enzymatic deglycosylation in the presence of detergents (Lifson *et al.*, 1986; Matthews *et al.*, 1987) have concluded that the carbohydrates are not directly involved with the binding, but that they are required to maintain the conformation of gp120 necessary for binding. In contrast, Fenouillet *et al.* (1989) enzymatically deglycosylated gp120 without detergent and demonstrated that the CD4 binding affinity was preserved. It therefore appears that the carbohydrate moieties of gp120 are not required for its binding to CD4 but that the conformational stability of gp120 to detergents is lost after deglycosylation.

The rgp120 used for these determinations is functionally and structurally equivalent to gp120 produced by HIV-1 infected cells.

REFERENCES

Allan, J.S., Coligan, J.E., Barin, F., McLane, M.F., Sodroski, J.G., Rosen, C.A., Haseltine, W.A., Lee, T.H. and Essex, M. (1985). *Science* 228, 1091-1094.

Arthur, L.O., Pyle, S.W., Nara, P.L., Bess, J.W. Jr., Gonda, M.A., Kelliher, J.C., Gilden, R.V., Robey, W.G., Bolognesi, D.P., Gallo, R.C. and Fischinger, P.J. (1987). *Proc. Natl. Acad. Sci. USA* 84, 8583-8587.

Barre-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W. and Montagnier, L. (1983). *Science* 220, 868-871.

Berman, P.W., Groopman, J.E., Gregory, T.J., Weiss, R.A., Clapham, P.R., Ferriani, R., Riddle, L., Shimasaki, C., Lucas, C., Lasky, L.A. and Eichberg, J.W. (1988). *Proc. Natl. Acad. Sci. USA* 85, 5200-5204.

Carr, S.A. and Roberts, G.D. (1986). *Anal. Biochem.* 157, 396-406.

Chu, F. (1986). *J. Biol. Chem.* 261, 172-177.

Cordonnier, A., Montagnier, L. and Emerman, M. (1989). *Nature* 340, 571-574.

Dalgleish, A.G., Beverly, P.C.L., Clapham, P.R., Crawford, D.H., Greaves, M.F. and Weiss, R.A. (1984). *Nature* 312, 763-766.

Diamond, D.C., Sleckman, B.P., Gregory, T., Lasky, L.A., Greenstein, J.L. and Burakoff, S.J. (1988). *J. Immunol.* 141, 3715-3717.

Drapeau, G.R. (1980). *J. Biol. Chem.* 255, 839-840.

Drapeau, G.R., Boily, Y. and Houmard, J. (1972). *J. Biol. Chem.* 247, 6720-6726.

Ellman, G.L. (1959). *Arch. Biochem. Biophys.* 82, 70-77.

Fenouillet, E., Clerget-Raslain, B., Gluckman, J.C., Guetard, D., Montagnier, L. and Bahraoui, E. (1989). *J. Exp. Med.* **169**, 807-822.

Gelderblom, H.R., Reupke, H. and Pauli, G. (1985). *Lancet* ii, 1016-1017.

Gelderblom, H.R., Hausmann, E.H.S., Ozel, M., Pauli, G. and Koch, M.A. (1987). *Virol.* **156**, 171-176.

Geyer, H., Holschbach, C., Hunsmann, G. and Schneider, J. (1988). *J. Biol. Chem.* **263**, 11760-11767.

Goudsmit, J., Boucher, C.A.B., Meloen, R.H., Epstein, L.G., Smit, L., van der Hoek, L. and Bakker, M. (1988). *AIDS* **2**, 157-164.

Ingram, V. (1963). *Meth. Enzymol.* **6**, 831-848.

Javaherian, K., Langlois, A.J., McDanal, C., Ross, K., Eckler, L.I., Jellis, C.L., Profy, A.T., Rusche, J.R., Bolognesi, D.P., Putney, S.D. and Matthews, T.J. (1989). *Proc. Natl. Acad. Sci. USA* **86**, 6768-6772.

Klatzman, D., Champagne, E., Chamaret, S., Gruest, J., Guetard, D., Hercend, T., Gluckman, J.C. and Montagnier, L. (1984). *Nature* **312**, 767-768.

Kornfeld, R. and Kornfeld, S. (1985). *Annu. Rev. Biochem.* **54**, 631-664.

Kowalski, M., Potz, J., Basiripour, L., Dorfman, T., Goh, W.C., Terwilliger, E., Dayton, A., Rosen, C., Haseltine, W. and Sodroski, J. (1987). *Science* **237**, 1351-1355.

Lasky, L.A., Groopman, J.E., Fennie, C.W., Benz, P.M., Capon, D.J., Dowbenko, D.J., Nakamura, G.R., Nunes, W.M., Renz, M.E. and Berman, P.W. (1986). *Science* **233**, 209-212.

Lasky, L.A., Nakamura, G., Smith, D.H., Fennie, C., Shimasaki, C., Patzer, E., Berman, P., Gregory, T. and Capon, D.J. (1987). *Cell* **50**, 975-985.

Lifson, J., Coutre, S., Huang, E. and Engleman, E. (1986). *J. Exp. Med.* **164**, 2101-2106.

Matsushita, S., Robert-Guroff, M., Rusche, J., Koito, A., Hattori, T., Hoshino, H., Javaherian, K., Takatsuki, K. and Putney, S. (1988). *J. Virol.* **62**, 2107-2114.

Matthews, T.J., Weinhold, K.J., Lyerly, H.K., Langlois, A.J., Wigzell, H. and Bolognesi, D.P. (1987). *Proc. Natl. Acad. Sci. USA* **84**, 5424-5428.

Mizuochi, T., Spellman, M.W., Larkin, M., Solomon, J., Basa, L.J. and Feizi, T. (1988a). *Biochem. J.* **254**, 599-603.

Mizuochi, T., Spellman, M.W., Larkin, M., Solomon, J., Basa, L.J. and Feizi, T. (1988b). *Biomed. Chromatogr.* **2**, 260-270.

Modrow, S., Hahn, B.H., Shaw, G.M., Gallo, R.C., Wong-Stahl, F. and Wolf, H. (1987). *J. Virol.* **61**, 570-578.

Moore, J.P., McKeating, J.A., Jones, I.M., Stephens, P.E., Clements, G., Thomson, S. and Weiss, R.A. (1990). *AIDS Research*, in press.

Nygren, A., Bergman, T., Matthews, T., Jornvall, H. and Wigzell, H. (1988). *Proc. Natl. Acad. Sci. USA* **85**, 6543-6546.

Paxton, R.J., Mooser, G., Pande, H., Lee, T.D., Shively, J.E. (1987). *Proc. Natl. Acad. Sci. USA* 84, 920-924.

Robey, W.G., Safai, B., Oroszlan, S., Arthur, L.O., Gonda, M.A., Gallo, R.C. and Fischinger, P.J. (1985). *Science* 228, 593-595.

Rusche, J.R., Javaherian, K., McDanal, C., Petro, J., Lynn, D.L., Grimaila, R., Langlois, A., Gallo, R.C., Arthur, L.O., Fischinger, P.J., Bolognesi, D.P., Putney, S.D. and Matthews, T.J. (1988). *Proc. Natl. Acad. Sci. USA* 85, 3198-3202.

Ryle, A.P. and Sanger, F. (1955). *Biochem. J.* 60, 535-540.

Sanger, F. and Thompson, E.O.P. (1953). *Biochem. J.* 53, 366-374.

Shalaby, M.R., Krowka, J.F., Gregory, T.J., Hirabayashi, S.E., McCabe, S.M., Kaufman, D.S., Stites, D.P. and Ammann, A.J. (1987). *Cell. Immunol.* 110, 140-148.

Siliciano, R.F., Lawton, T., Knall, C., Karr, R.W., Berman, P., Gregory, T. and Reinherz, E.L. (1988). *Cell* 54, 561-575.

Sodroski, J., Goh, W.C., Rosen, C., Campbell, K. and Haseltine, W.A. (1986). *Nature* 322, 470-474.

Tai, T., Yamashita, K. and Kobata, A. (1977). *Biochem. Biophys. Res. Commun.* 78, 434-441.

Tarentino, A.L., Gomez, C.M. and Plummer, T.H. Jr. (1985). *Biochemistry* 24, 4665-4671.

Tarentino, A.L., Plummer, T.H. Jr. and Maley, F. (1974). *J. Biol. Chem.* 249, 818-824.

Willey, R.L., Rutledge, R.A., Dias, S., Folks, T., Theodore, T., Buckler, C.E. and Martin, M.A. (1986). *Proc. Natl. Acad. Sci. USA* 83, 5038-5042.

Willey, R.L., Smith, D.H., Lasky, L.A., Theodore, T.S., Earl, P.L., Moss, B., Capon, D.J. and Martin, M.A. (1988). *J. Virol.* 62, 139-147.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Genentech, Inc.

(ii) TITLE OF INVENTION: HIV Envelope Polypeptides

(iii) NUMBER OF SEQUENCES: 15

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Genentech, Inc.
(B) STREET: 460 Point San Bruno Blvd
(C) CITY: South San Francisco
(D) STATE: California
(E) COUNTRY: USA
(F) ZIP: 94080

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: patin (Genentech)

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: U.S.S.N. 07/504,772
(B) FILING DATE: 03-APRIL-1990

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Adler, Carolyn R.
(B) REGISTRATION NUMBER: 32,324
(C) REFERENCE/DOCKET NUMBER: 639

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 415/266-2614
(B) TELEFAX: 415/952-9881
(C) TELEX: 910/371-7168

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

Cys Val Lys Leu Thr Pro Leu Cys Cys Asn Thr Ser Val Ile Thr
1 5 10 15

Gln Ala Cys
18

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 40 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Pro Ile His Tyr Cys Ala Pro Ala Gly Phe Ala Ile Leu Lys Cys
1 5 10 15

Asn Asn Lys Thr Phe Asn Gly Thr Gly Pro Cys Thr Asn Val Ser
20 25 30

Thr Val Gln Cys Thr His Gly Ile Arg Pro
35 40

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Cys Asn Asn Lys Thr Phe Asn Gly Thr Gly Pro Cys
1 5 10 12

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Cys Ala Pro Ala Gly Phe Ala Ile Leu Lys Cys Cys Thr Asn Val
1 5 10 15

Ser Thr Val Gln Cys
20

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

Pro Ile His Tyr Cys Cys Thr His Gly Ile Arg Pro
1 5 10 12

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 58 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Gly Gly Asp Pro Glu Ile Val Thr His Ser Phe Asn Cys Gly Gly
1 5 10 15

Glu Phe Phe Tyr Cys Asn Ser Leu Pro Cys Arg Ile Lys Gln Phe
20 25 30

Ile Asn Met Trp Gln Glu Val Gly Lys Ala Met Tyr Ala Pro Pro
35 40 45

Ile Ser Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr Gly
50 55 58

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Cys Gly Gly Glu Phe Phe Tyr Cys Cys Arg Ile Lys Gln Phe Ile
1 5 10 15

Asn Met Trp Gln Glu Val Gly Lys Ala Met Tyr Ala Pro Pro Ile
20 25 30

Ser Gly Gln Ile Arg Cys
35 36

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr Glu Val His Asn Val
1 5 10 15

Trp Ala Thr His Ala Cys
20 21

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala Tyr Asp Thr
1 5 10 15

Glu Val His Asn Val Trp Ala Thr His Ala Cys Val Pro Thr Asp
20 25 30

Pro Asn
32

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 479 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Thr Glu Lys Leu Trp Val Thr Val Tyr Tyr Gly Val Pro Val Trp
1 5 10 15

Lys Glu Ala Thr Thr Thr Leu Phe Cys Ala Ser Asp Ala Lys Ala
20 25 30

Tyr Asp Thr Glu Val His Asn Val Trp Ala Thr His Ala Cys Val
35 40 45

Pro Thr Asp Pro Asn Pro Gln Glu Val Val Leu Val Asn Val Thr
50 55 60

Glu Asn Phe Asn Met Trp Lys Asn Asp Met Val Glu Gln Met His
65 70 75

Glu Asp Ile Ile Ser Leu Trp Asp Gln Ser Leu Lys Pro Cys Val
80 85 90

Lys Leu Thr Pro Leu Cys Val Ser Leu Lys Cys Thr Asp Leu Lys
95 100 105

Asn Asp Thr Asn Thr Asn Ser Ser Ser Gly Arg Met Ile Met Glu
110 115 120

Lys Gly Glu Ile Lys Asn Cys Ser Phe Asn Ile Ser Thr Ser Ile
125 130 135

Arg Gly Lys Val Gln Lys Glu Tyr Ala Phe Phe Tyr Lys Leu Asp
140 145 150

Ile Ile Pro Ile Asp Asn Asp Thr Thr Ser Tyr Thr Leu Thr Ser

155 160 165

Cys Asn Thr Ser Val Ile Thr Gln Ala Cys Pro Lys Val Ser Phe
170 175 180

Glu Pro Ile Pro Ile His Tyr Cys Ala Pro Ala Gly Phe Ala Ile
185 190 195

Leu Lys Cys Asn Asn Lys Thr Phe Asn Gly Thr Gly Pro Cys Thr
200 205 210

Asn Val Ser Thr Val Gln Cys Thr His Gly Ile Arg Pro Val Val
215 220 225

Ser Thr Gln Leu Leu Leu Asn Gly Ser Leu Ala Glu Glu Glu Val
230 235 240

Val Ile Arg Ser Ala Asn Phe Thr Asp Asn Ala Lys Thr Ile Ile
245 250 255

Val Gln Leu Asn Gln Ser Val Glu Ile Asn Cys Thr Arg Pro Asn
260 265 270

Asn Asn Thr Arg Lys Ser Ile Arg Ile Gln Arg Gly Pro Gly Arg
275 280 285

Ala Phe Val Thr Ile Gly Lys Ile Gly Asn Met Arg Gln Ala His
290 295 300

Cys Asn Ile Ser Arg Ala Lys Trp Asn Asn Thr Leu Lys Gln Ile
305 310 315

Asp Ser Lys Leu Arg Glu Gln Phe Gly Asn Asn Lys Thr Ile Ile
320 325 330

Phe Lys Gln Ser Ser Gly Gly Asp Pro Glu Ile Val Thr His Ser
335 340 345

Phe Asn Cys Gly Gly Glu Phe Phe Tyr Cys Asn Ser Thr Gln Leu
350 355 360

Phe Asn Ser Thr Trp Phe Asn Ser Thr Trp Ser Thr Glu Gly Ser
365 370 375

Asn Asn Thr Glu Gly Ser Asp Thr Ile Thr Leu Pro Cys Arg Ile
380 385 390

Lys Gln Phe Ile Asn Met Trp Gln Glu Val Gly Lys Ala Met Tyr
395 400 405

Ala Pro Pro Ile Ser Gly Gln Ile Arg Cys Ser Ser Asn Ile Thr
410 415 420

Gly Leu Leu Leu Thr Arg Asp Gly Gly Asn Asn Asn Asn Glu Ser
425 430 435

Glu Ile Phe Arg Pro Gly Gly Gly Asp Met Arg Asp Asn Trp Arg
440 445 450

Ser Glu Leu Tyr Lys Tyr Lys Val Val Lys Ile Glu Pro Leu Gly
455 460 465

Val Ala Pro Thr Lys Ala Lys Arg Arg Val Val Gln Arg Glu
470 475 479

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 9 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Lys Tyr Ala Leu Ala Asp Ala Ser Leu
1 5 9

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

Lys Tyr Ala Leu Ala Asp Ala Ser Leu Lys Met Ala Asp Pro Asn
1 5 10 15

Arg Phe Arg Gly Lys Asp Leu Pro Val Leu Asp Gln
20 25 27

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 481 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

Thr Gln Tyr Val Thr Val Phe Tyr Gly Val Pro Thr Trp Lys Asn
1 5 10 15

Ala Thr Ile Pro Leu Phe Cys Ala Thr Arg Asn Arg Asp Thr Trp
20 25 30

Gly Thr Ile Gln Cys Leu Pro Asp Asn Asp Asp Tyr Gln Glu Ile
35 40 45

Thr Leu Asn Val Thr Glu Ala Phe Asp Ala Trp Asn Asn Thr Val
50 55 60

```
Thr Glu Gln Ala Ile Glu Asp Val Trp His Leu Phe Glu Thr Ser
65                  70                  75

Ile Lys Pro Cys Val Lys Leu Thr Pro Leu Cys Val Ala Met Lys
80                  85                  90

Cys Ser Ser Thr Glu Ser Ser Thr Gly Asn Asn Thr Thr Ser Lys
95                  100                 105

Ser Thr Ser Thr Thr Thr Thr Thr Pro Thr Asp Gln Glu Gln Glu
110                 115                 120

Ile Ser Glu Asp Thr Pro Cys Ala Arg Ala Asp Asn Cys Ser Gly
125                 130                 135

Leu Gly Glu Glu Glu Thr Ile Asn Cys Gln Phe Asn Met Thr Gly
140                 145                 150

Leu Glu Arg Asp Lys Lys Lys Gln Tyr Asn Glu Thr Trp Tyr Ser
155                 160                 165

Lys Asp Val Val Cys Glu Thr Asn Asn Ser Thr Asn Gln Thr Gln
170                 175                 180

Cys Tyr Met Asn His Cys Asn Thr Ser Val Ile Thr Glu Ser Cys
185                 190                 195

Asp Lys His Tyr Trp Asp Ala Ile Arg Phe Arg Tyr Cys Ala Pro
200                 205                 210

Pro Gly Tyr Ala Leu Leu Arg Cys Asn Asp Thr Asn Tyr Ser Gly
215                 220                 225

Phe Ala Pro Asn Cys Ser Lys Val Val Ala Ser Thr Cys Thr Arg
230                 235                 240

Met Met Glu Thr Gln Thr Ser Thr Trp Phe Gly Phe Asn Gly Thr
245                 250                 255

Arg Ala Glu Asn Arg Thr Tyr Ile Tyr Trp His Gly Arg Asp Asn
260                 265                 270

Arg Thr Ile Ile Ser Leu Asn Lys Tyr Tyr Asn Leu Ser Leu His
275                 280                 285

Cys Lys Arg Pro Gly Asn Lys Ile Val Lys Gln Ile Met Leu Met
290                 295                 300

Ser Gly His Val Phe His Ser His Gln Pro Ile Asn Lys Arg Pro
305                 310                 315

Arg Gln Ala Trp Cys Trp Phe Lys Gly Lys Trp Lys Asp Ala Met
320                 325                 330

Gln Glu Val Lys Glu Thr Leu Ala Lys His Pro Arg Tyr Arg Gly
335                 340                 345
```

Thr Asn Asp Thr Arg Asn Ile Ser Phe Ala Ala Pro Gly Lys Gly
350 355 360

Ser Asp Pro Glu Val Ala Tyr Met Trp Thr Asn Cys Arg Gly Glu
365 370 375

Phe Leu Tyr Cys Asn Met Thr Trp Phe Leu Asn Trp Ile Glu Asn
380 385 390

Lys Thr His Arg Asn Tyr Ala Pro Cys His Ile Lys Gln Ile Ile
395 400 405

Asn Thr Trp His Lys Val Gly Arg Asn Val Tyr Leu Pro Pro Arg
410 415 420

Glu Gly Glu Leu Ser Cys Asn Ser Thr Val Thr Ser Ile Ile Ala
425 430 435

Asn Ile Asp Trp Gln Asn Asn Asn Gln Thr Asn Ile Thr Phe Ser
440 445 450

Ala Glu Val Ala Glu Leu Tyr Arg Leu Glu Leu Gly Asp Tyr Lys
455 460 465

Leu Val Glu Ile Thr Pro Ile Gly Phe Ala Pro Thr Lys Glu Lys
470 475 480

Arg
481

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 8 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

Gln Ala His Cys Asn Ile Ser Arg
1 5 8

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Cys Asn Asn Lys
1 4

## Claims

We claim:

1. An isolated cyclized polypeptide sequence comprising the amino acid residues selected from the group consisting of:
   a) C V K L T P L C C N T S V I T Q A C [SEQ. ID NO. 1] and containing less than about 28 amino acid residues;
   b) P I H Y C A P A G F A I L K C N N K T F N G T G P C T N V S T V Q C T H G I R P [SEQ. ID NO. 2] and containing less than about 45 amino acid residues;
   c) C N N K T F N G T G P C [SEQ. ID NO. 3] and containing less than about 22 amino acid residues;
   d) C A P A G F A I L K C C T N V S T V Q C [SEQ. ID NO. 4] and containing less than about 30 amino acid residues;
   e) P I H Y C C T H G I R P [SEQ. ID NO. 5] and containing less than about 22 amino acid residues;
   f) G G D P E I V T H S F N C G G E F F Y C N S L P C R I K Q F I N M W Q E V G K A M Y A P P I S G Q I R C S S N I T G [SEQ. ID NO. 6] and containing less than about 65 amino acid residues;
   g) C G G E F F Y C C R I K Q F I N M W Q E V G K A M Y A P P I S G Q I R C [SEQ. ID NO. 7] and containing less than about 45 amino acid residues;
   h) C A S D A K A Y D T E V H N V W A T H A C [SEQ. ID NO. 8] and containing less than about 30 amino acid residues; and
   i) T T T L F C A S D A K A Y D T E V H N V W A T H A C V P T D P N [SEQ. ID NO. 9] and containing less than about 50 amino acid residues.
2. A method for the prophylaxis or treatment of HIV infection comprising administering a therapeutically effective dose of a sterile composition comprising the cyclized peptide of claim 1 and an pharmaceutically acceptable vehicle to a patient having or at risk of having HIV infection.
3. The method of claim 2 wherein the therapeutic dose is about from $0.5 \times 10^{-6}$ to $5 \times 10^{-6}$ molar.
4. The method of claim 2 wherein the composition further contains an adjuvant.
5. An antibody which is directed to an antigenic determinant comprised by the isolated cyclized polypeptide of claim 1.
6. The antibody of claim 5 which is conjugated to a cytotoxin.
7. The antibody of claim 5 which is covalently bound to a detectable marker or a water-insoluble matrix.
8. The antibody of claim 5 in a sterile, pharmaceutically acceptable vehicle.

9. An isolated polypeptide having an antigenic determinant or determinants immunologically cross-reactive with a determinant of an HIV env polypeptide having an amino acid sequence selected from the group consisting of

a) residues 1-80;

b) residues 8-180;

c) residues 165-260;

d) residues 160-260;

e) residues 260-310; and

f) residues 320-479.

10. An antibody directed to an isolated polypeptide having an antigenic determinant or determinants immunologically cross-reactive with a determinant of the HIV env polypeptide of strain HTLV-IIIB having an amino acid sequence selected from the group consisting of:

a) residues 1-80;

b) residues 8-180;

c) residues 165-260;

d) residues 160-260;

e) residues 260-310; and

f) residues 320-479.

11. The antibody of claim 10 which is conjugated to a cytotoxin.

12. The antibody of claim 10 which is covalently bound to a detectable marker or a water-insoluble matrix.

13. The antibody of claim 10 in a sterile, pharmaceutically acceptable vehicle.

14. A method for the prophylaxis or treatment of HIV infection comprising administering a therapeutically effective dose of a sterile composition comprising the antibody of claim 5 and an pharmaceutically acceptable vehicle to a patient having or at risk of having HIV infection.

15. The method of claim 14, wherein said antibody is conjugated to a cytotoxin.

16. A method for the prophylaxis or treatment of HIV infection comprising administering a therapeutically effective dose of a sterile composition comprising the antibody of claim 10 and an pharmaceutically acceptable vehicle to a patient having or at risk of having HIV infection.

17. The method of claim 16, wherein said antibody is conjugated to a cytotoxin.

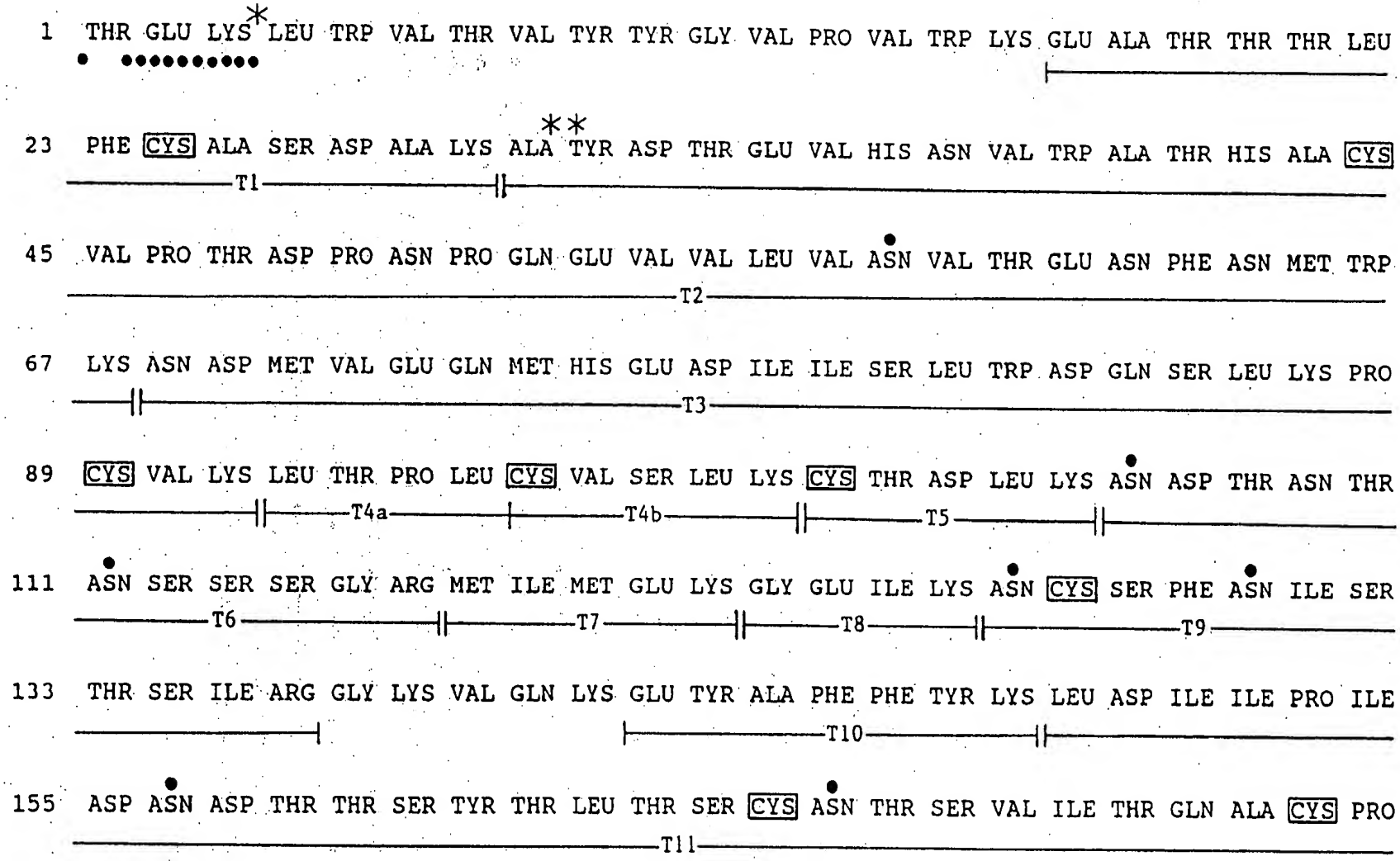

1 THR GLU LYS LEU TRP VAL THR VAL TYR TYR GLY VAL PRO VAL TRP LYS GLU ALA THR THR THR LEU
23 PHE CYS ALA SER ASP ALA LYS ALA TYR ASP THR GLU VAL HIS ASN VAL TRP ALA THR HIS ALA CYS
T1
45 VAL PRO THR ASP PRO ASN PRO GLN GLU VAL VAL LEU VAL ASN VAL THR GLU ASN PHE ASN MET TRP
T2
67 LYS ASN ASP MET VAL GLU GLN MET HIS GLU ASP ILE ILE SER LEU TRP ASP GLN SER LEU LYS PRO
T3
89 CYS VAL LYS LEU THR PRO LEU CYS VAL SER LEU LYS CYS THR ASP LEU LYS ASN ASP THR ASN THR
T4a
T4b
T5
111 ASN SER SER SER GLY ARG MET ILE MET GLU LYS GLY GLU ILE LYS ASN CYS SER PHE ASN ILE SER
T6
T7
T8
T9
133 THR SER ILE ARG GLY LYS VAL GLN LYS GLU TYR ALA PHE PHE TYR LYS LEU ASP ILE ILE PRO ILE
T10
155 ASP ASN ASP THR THR SER TYR THR LEU THR SER CYS ASN THR SER VAL ILE THR GLN ALA CYS PRO
T11

FIG. 1A-1

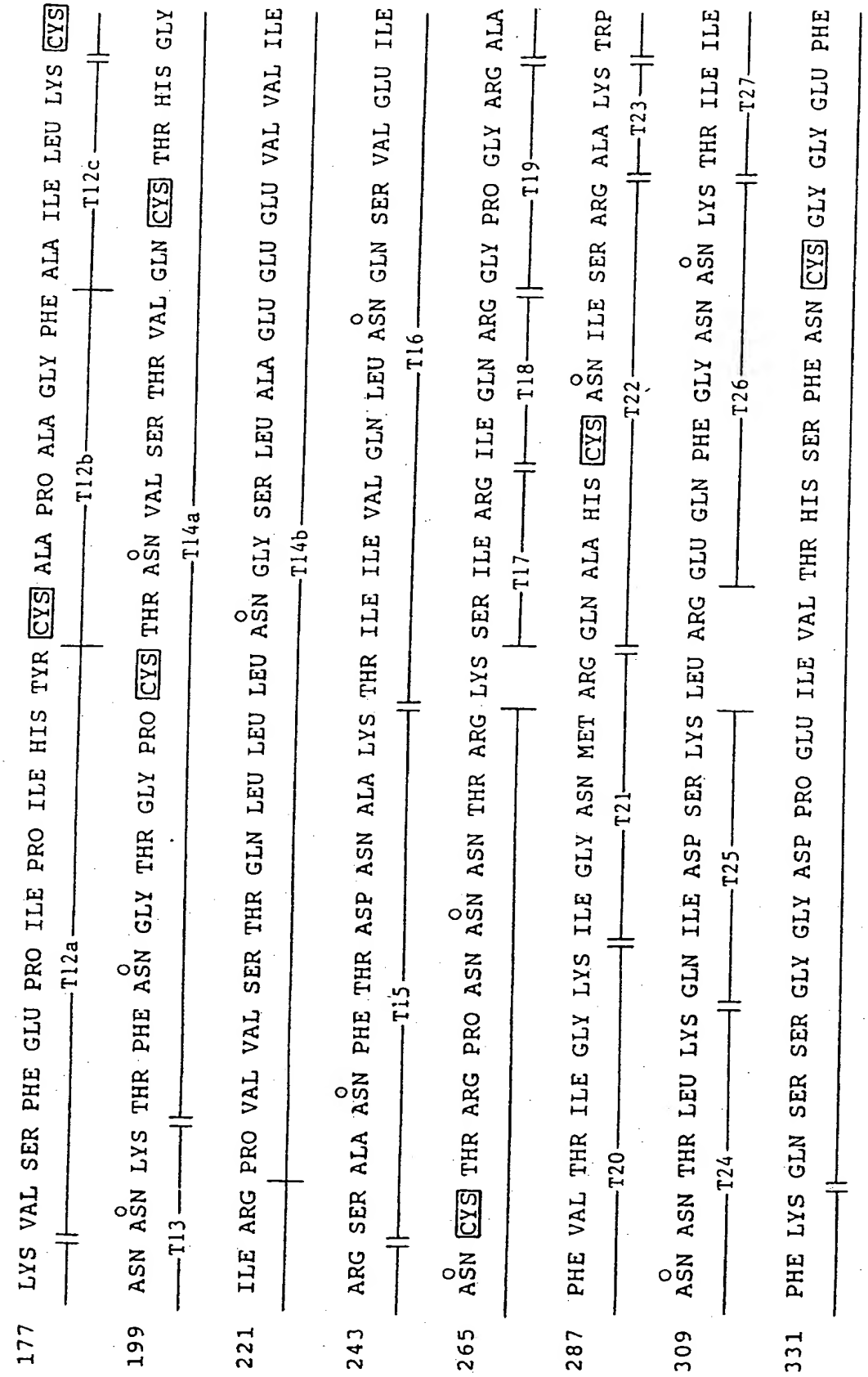

177 LYS VAL SER PHE GLU PRO ILE PRO ILE HIS TYR CYS ALA PRO ALA GLY PHE ALA ILE LEU LYS CYS
T12a
T12b
T12c
199 ASN ASN LYS THR PHE ASN GLY THR GLY PRO CYS THR ASN VAL SER THR VAL GLN CYS THR HIS GLY
T13
T14a
221 ILE ARG PRO VAL VAL SER THR GLN LEU LEU LEU ASN GLY SER LEU ALA GLU GLU GLU VAL VAL ILE
T14b
243 ARG SER ALA ASN PHE THR ASP ASN ALA LYS THR ILE ILE VAL GLN LEU ASN GLN SER VAL GLU ILE
T15
T16
265 ASN CYS THR ARG PRO ASN ASN THR ARG LYS SER ILE ARG ILE GLN ARG GLY PRO GLY ARG ALA
T17
T18
T19
287 PHE VAL THR ILE GLY LYS ILE GLY ASN MET ARG GLN ALA HIS CYS ASN ILE SER ARG ALA LYS TRP
T20
T21
T22
T23
309 ASN ASN THR LEU LYS GLN ILE ASP SER LYS LEU ARG GLU GLN PHE GLY ASN ASN LYS THR ILE ILE
T24
T25
T26
T27
331 PHE LYS GLN SER SER GLY GLY ASP PRO GLU ILE VAL THR HIS SER PHE ASN CYS GLY GLY GLU PHE

FIG. 1A-2

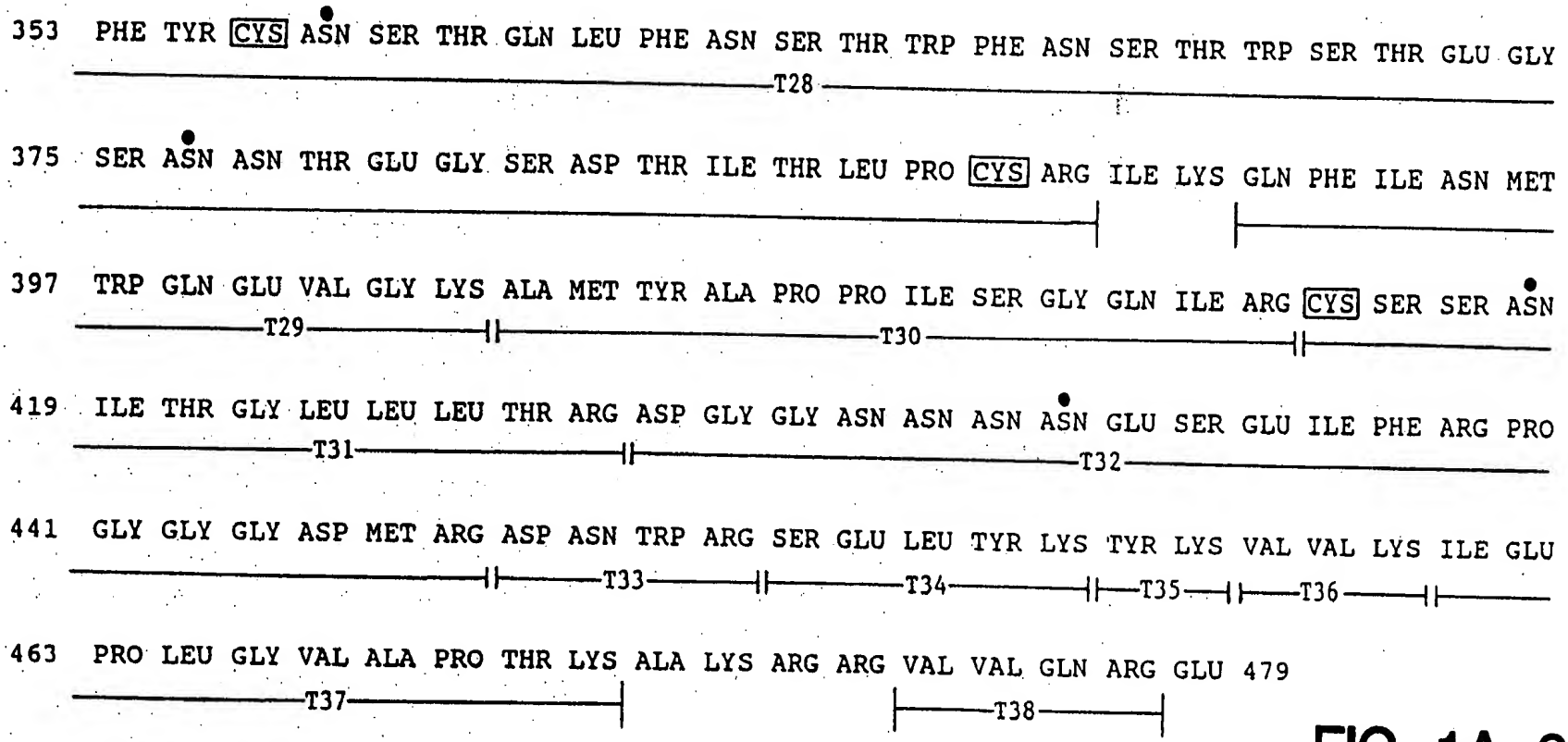



FIG. 1A-3

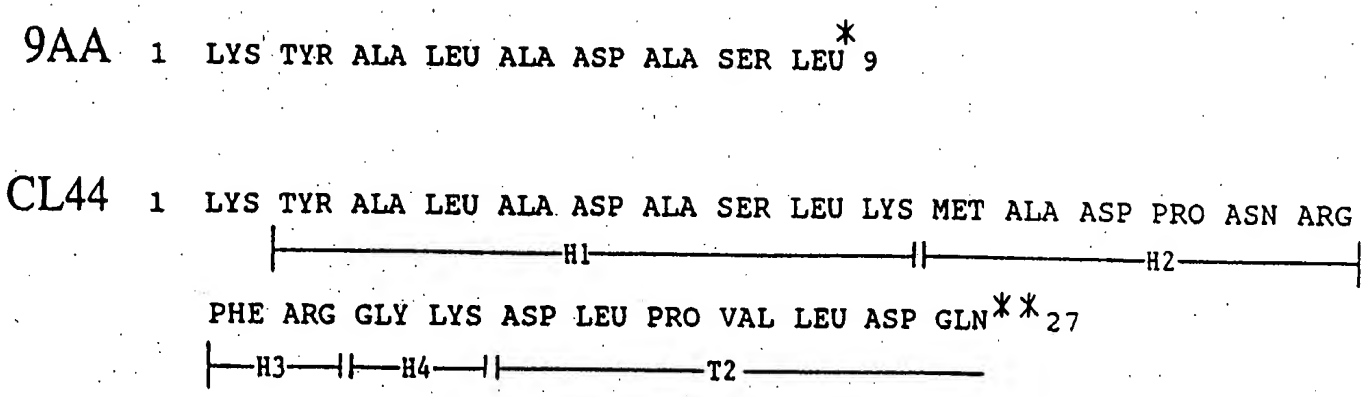



FIG. 1B

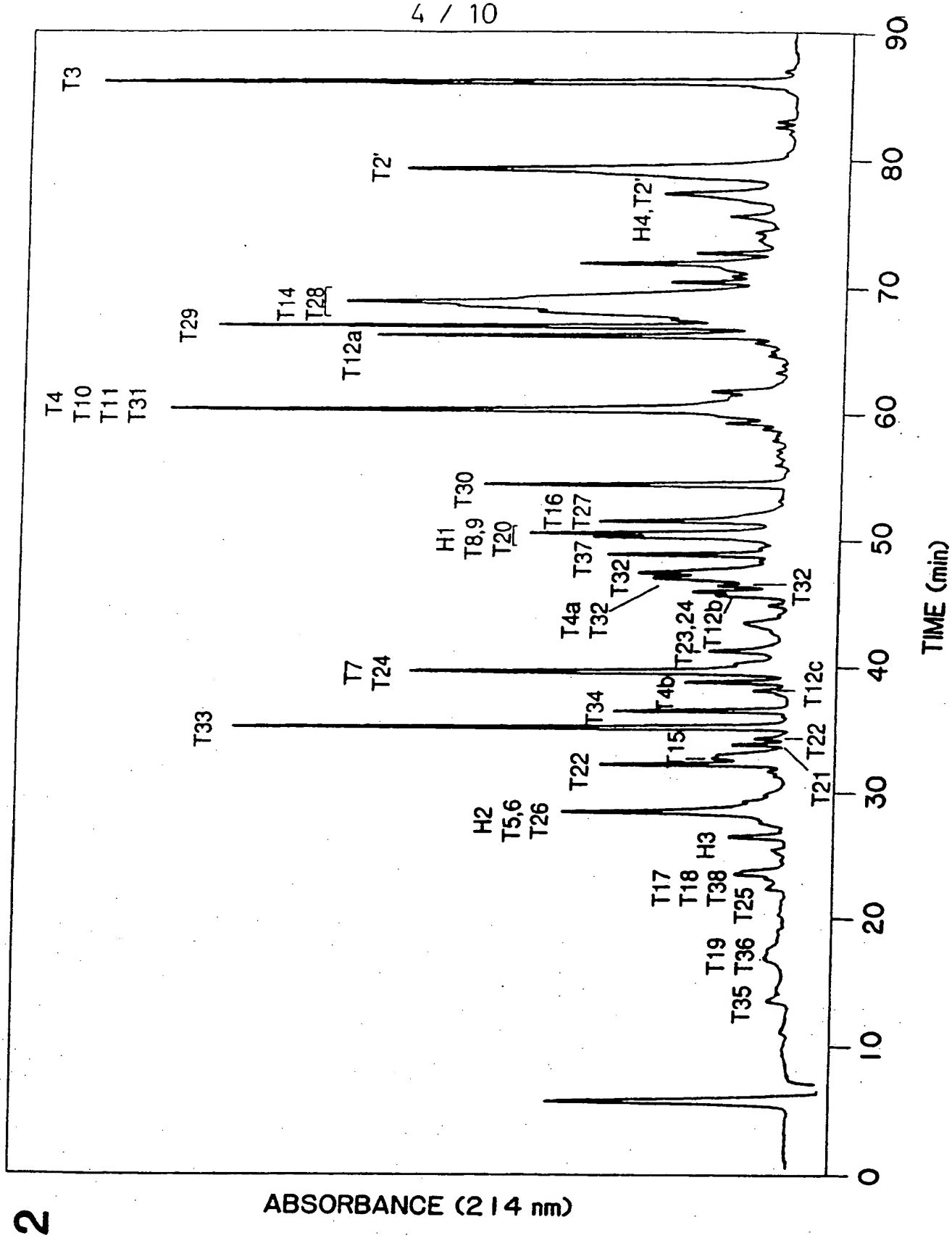

T3
T2'
H4,T2'
T14
T28
T29
T12a
T4
T10
T11
T31
H1
T8,9
T20
T30
T16
T27
T37
T32
T4a
T32
T32
T12b
T23,24
T7
T24
T4b
T12c
T34
T33
T15
T22
T21
T22
H2
T5,6
T26
H3
T17
T18
T38
T25
T19
T36
T35
0
10
20
30
40
50
60
70
80
90
TIME (min)
ABSORBANCE (214 nm)

FIG. 2

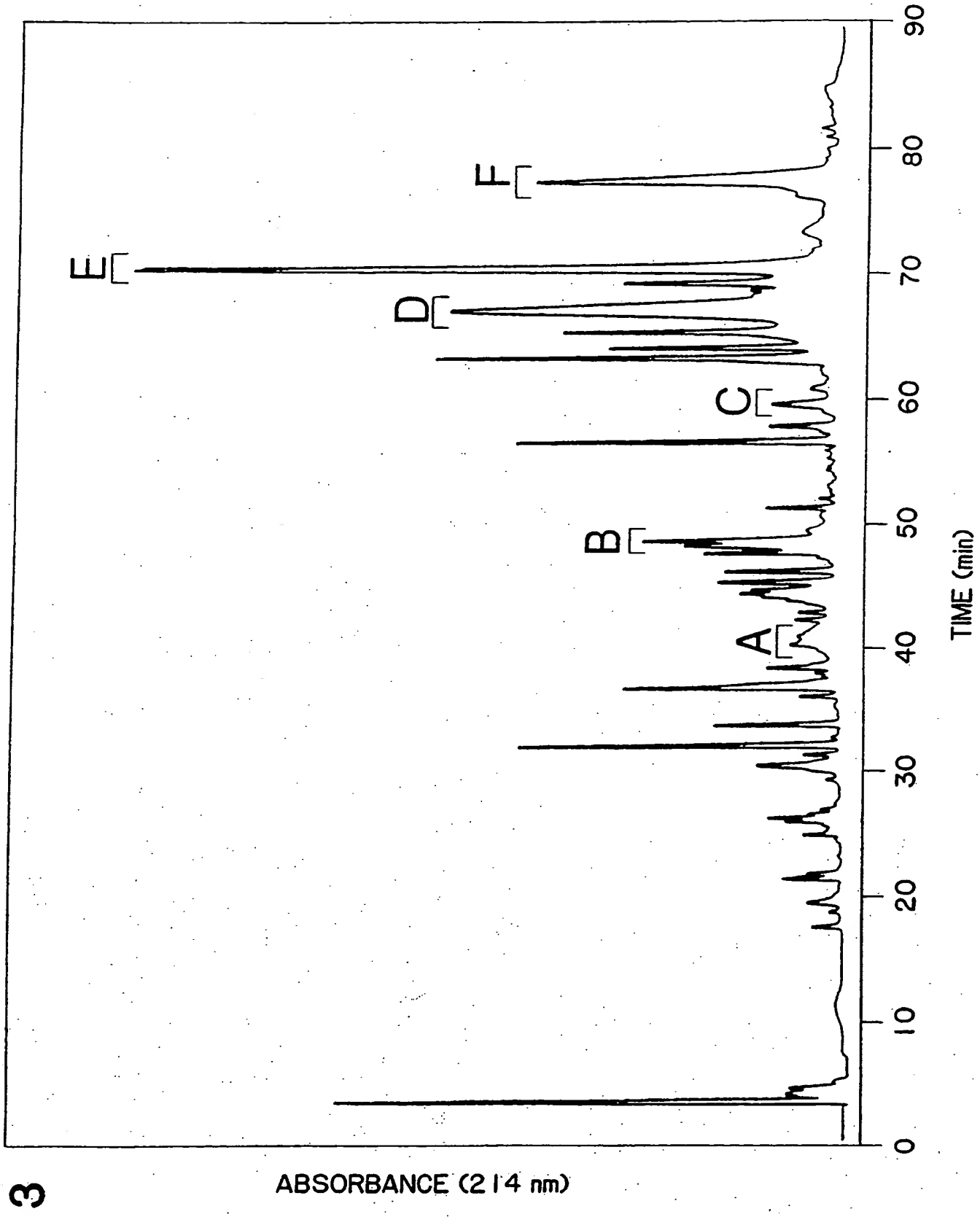



FIG. 3

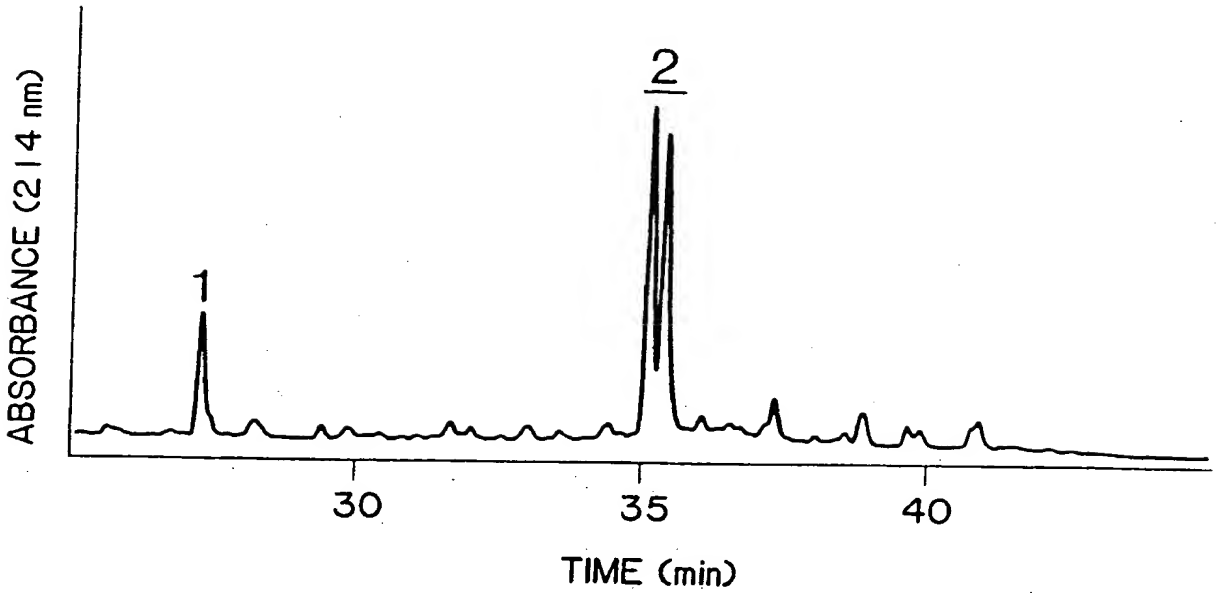

ABSORBANCE (214 nm)
1
2
30
35
40
TIME (min)

FIG. 4A

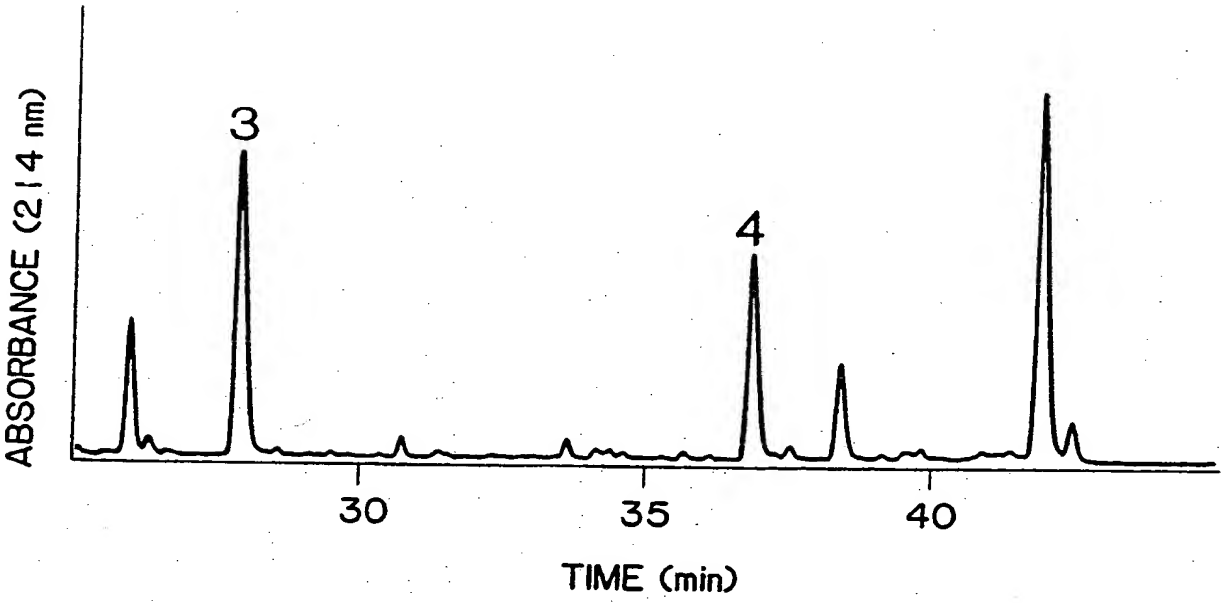

ABSORBANCE (214 nm)
3
4
30
35
40
TIME (min)

FIG. 4B

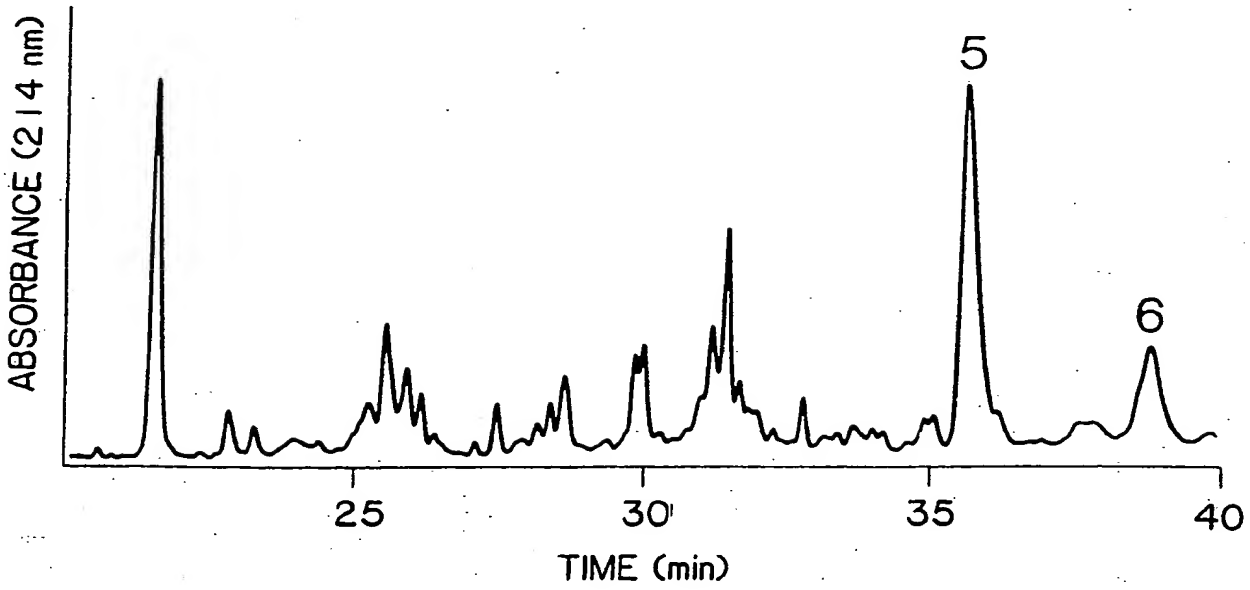

ABSORBANCE (214 nm)
5
6
25
30
35
40
TIME (min)

FIG. 4C

FIG. 5A

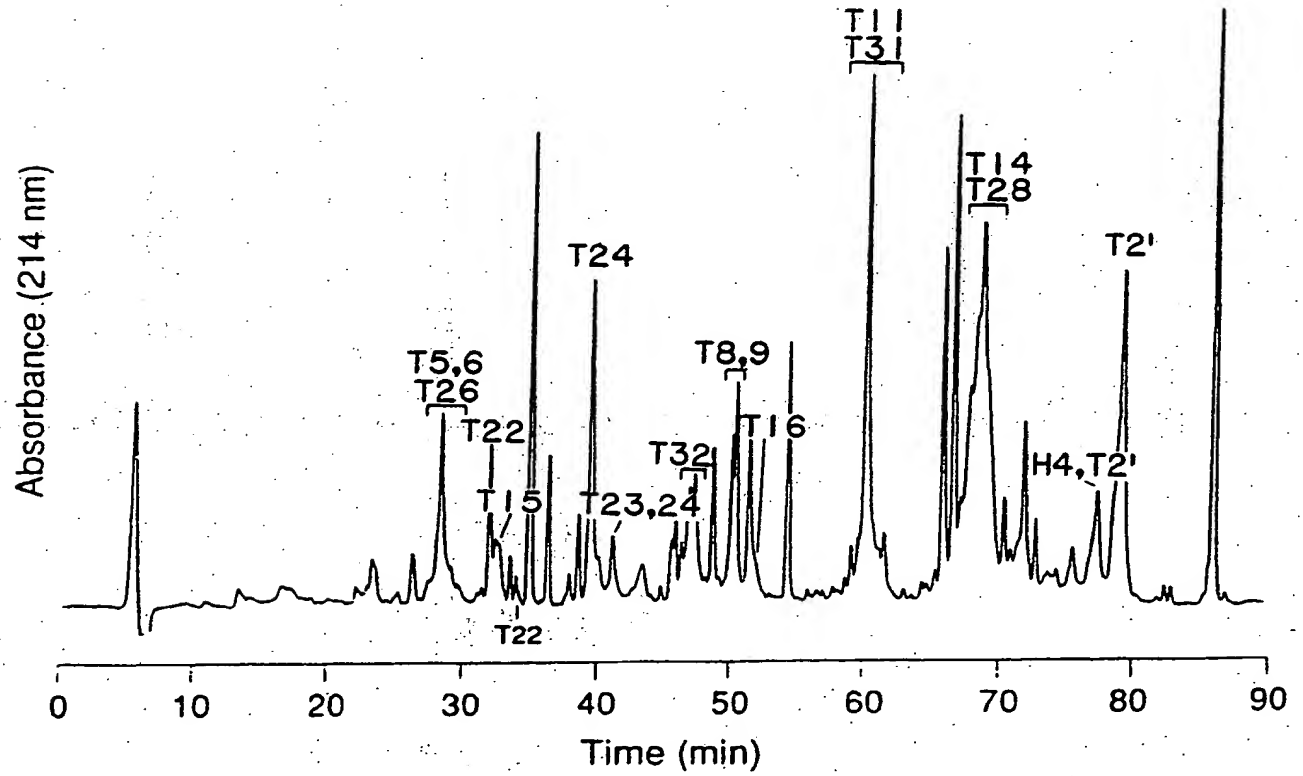

Absorbance (214 nm)
T11
T31
T14
T28
T24
T2'
T5,6
T26
T8,9
T22
T16
T32
T15
T23,24
H4,T2'
T22
0
10
20
30
40
50
60
70
80
90
Time (min)

FIG. 5B

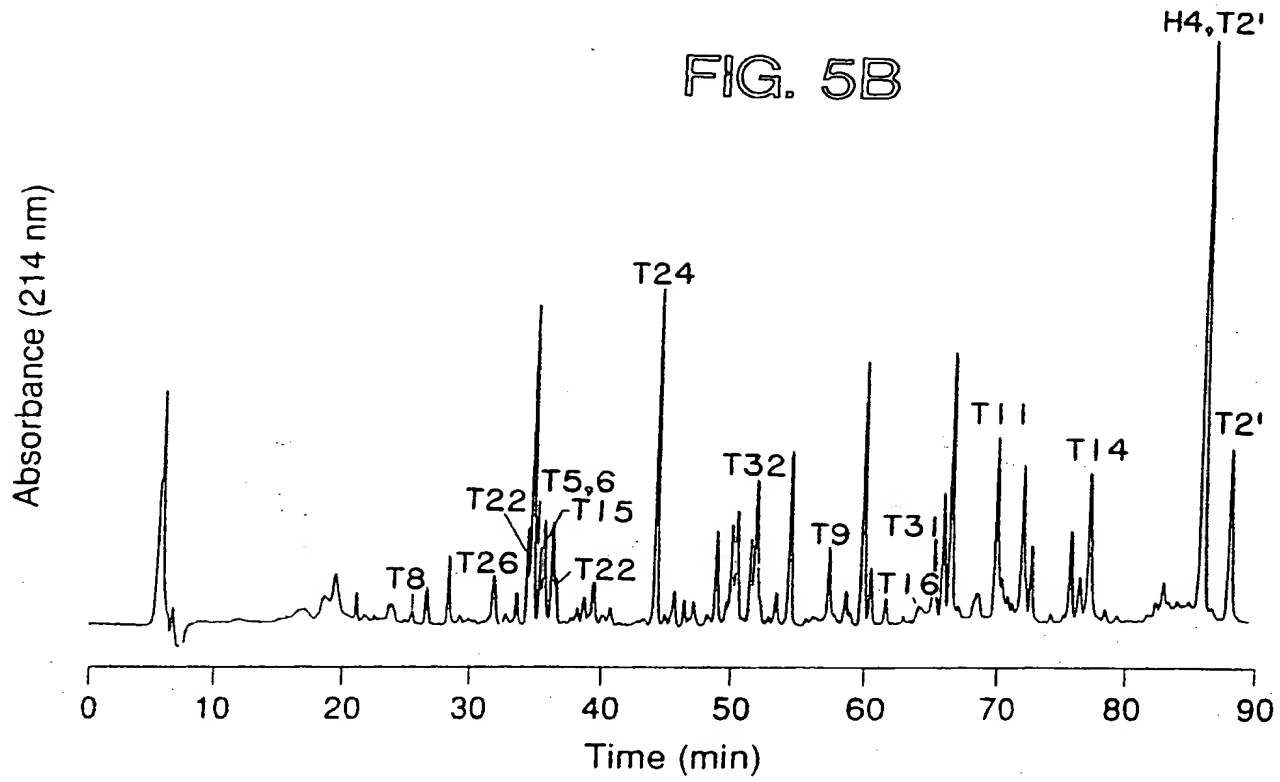

H4,T2'
T24
T11
T14
T2'
T32
T5,6
T22
T15
T9
T31
T26
T22
T8
T16
Absorbance (214 nm)
0
10
20
30
40
50
60
70
80
90
Time (min)

FIG. 5C

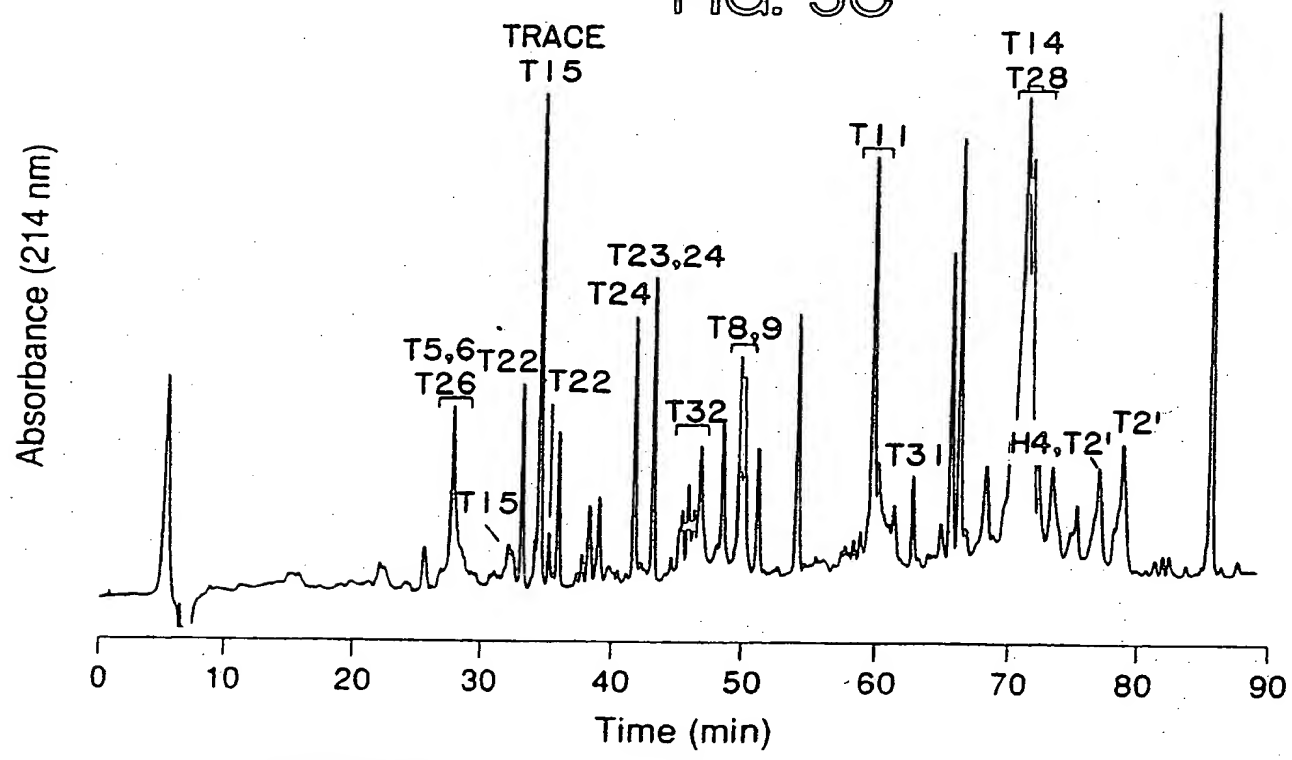
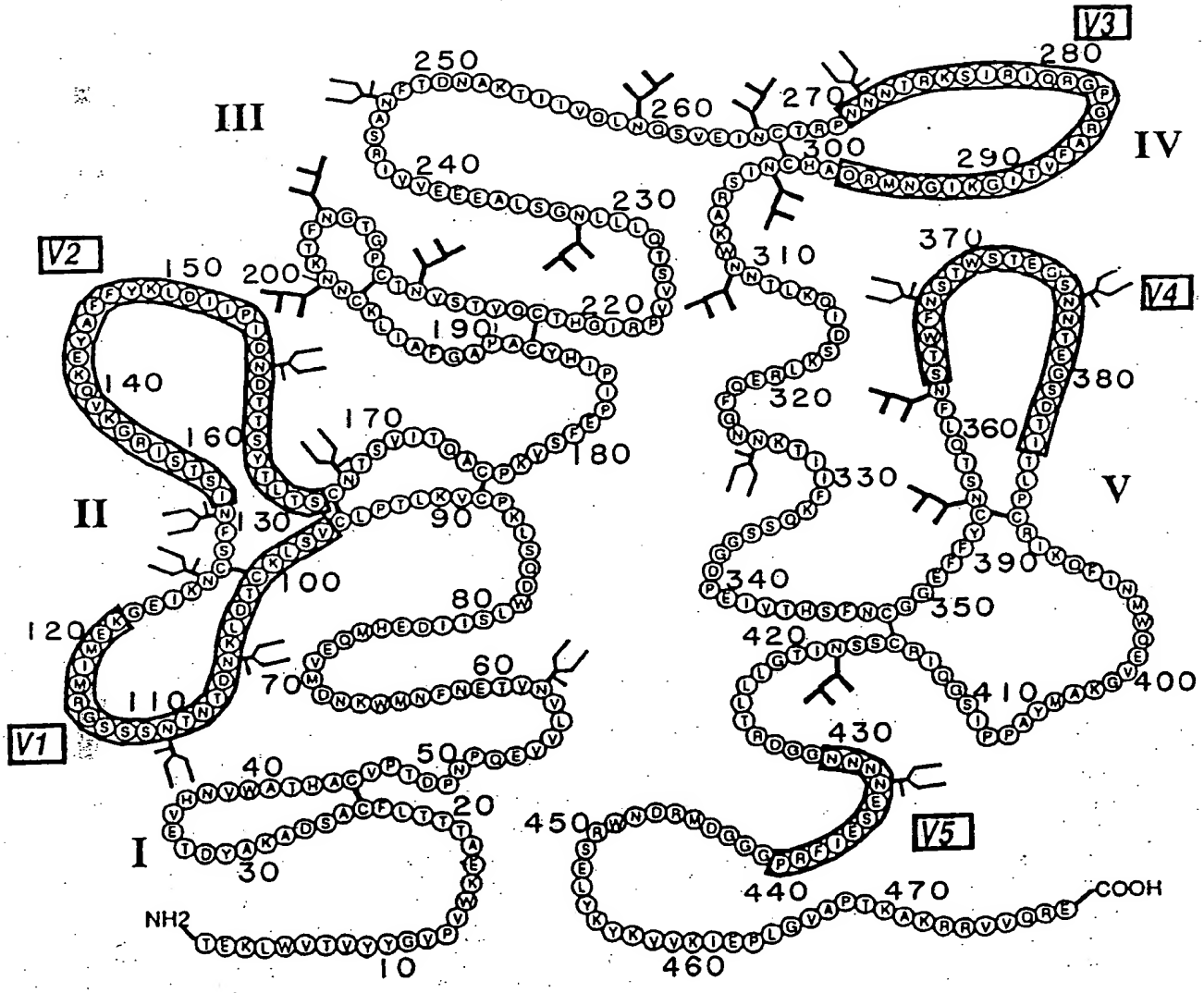

TRACE
T15
T14
T28
T11
T23,24
T24
T8,9
T5,6
T26
T22
T22
T32
T31
H4,T2'
T2'
T15
Absorbance (214 nm)
0
10
20
30
40
50
60
70
80
90
Time (min)

FIG. 6



HIV-2

FIG. 7